Title	Asymptotic Theory on the Least Squares Estimation of Threshold Moving-Average Models
Author(s)	Li, D; Ling, S; Li, WK
Citation	Econometric Theory, 2013, v. 29 n. 3, p. 482-516
Issued Date	2013
URL	http://hdl.handle.net/10722/191502
Rights	Econometric Theory. Copyright © Cambridge University Press.

Econometric Theory, **29**, 2013, 482–516.
doi:10.1017/S026646661200045X

ASYMPTOTIC THEORY ON THE LEAST SQUARES ESTIMATION OF THRESHOLD MOVING-AVERAGE MODELS

DONG LI AND SHIQING LING
Hong Kong University of Science and Technology

WAI KEUNG LI
University of Hong Kong

This paper studies the asymptotic theory of least squares estimation in a threshold moving average model. Under some mild conditions, it is shown that the estimator of the threshold is n-consistent and its limiting distribution is related to a two-sided compound Poisson process, whereas the estimators of other coefficients are strongly consistent and asymptotically normal. This paper also provides a resampling method to tabulate the limiting distribution of the estimated threshold in practice, which is the first successful effort in this direction. This resampling method contributes to threshold literature. Simultaneously, simulation studies are carried out to assess the performance of least squares estimation in finite samples.

1. INTRODUCTION

Since the threshold model was introduced by Tong (1978), it has become a more or less standard model in nonlinear time series. One of the leading reasons is that piecewise linear functions can offer a relatively simple and easy-to-handle approximation to the complex nonlinear dynamics. Threshold autoregressive (TAR) or TAR-type models have been widely used to study nonlinear phenomena in various fields; see Hansen (1997, 1999, 2000), So, Li, and Lam (2002), Tiao and Tsay (1994) in economics; Li and Lam (1995), Li and Li (1996), Liu, Li, and Li (1997), Yadav, Pope, and Paudyal (1994) in finance; Tong and Lim (1980) in hydrology; among others. The probabilistic structures of TAR-type models have been studied by many authors; see An and Huang (1996), Brockwell, Liu, and Tweedie (1992), Chan, Petruccelli, Tong, and Woolford (1985), Chan and Tong (1985), Chen and Tsay (1991), Cline and Pu (2004), Ling (1999), Liu and Susko (1992), Lu (1998),

S. Ling thanks Hong Kong Research Grants Council, grants HKUST602609 and HKUST641912, for partial support. W.K. Li thanks Hong Kong Research Grants Council, grant HKU7036/06P, for partial support. The authors also thank the co-editor and two anonymous referees for their helpful comments that improved the presentation. Address correspondence to Wai Keung Li, Department of Statistics and Actuarial Science, The University of Hong Kong, Pokfulam Road, Hong Kong; e-mail: hrntlwk@hku.hk.

and so on. The asymptotic theory of the least squares estimation (LSE) of the TAR model was established by Chan (1993) and Chan and Tsay (1998); see also Petruccelli (1986) and Qian (1998). The least absolute deviation estimation was considered by Caner (2002). An excellent review of the TAR model is available in Tong (1990), and a selective review of the history of threshold models is given by Tong (2011).

To date, however, most work in this area has been concentrated on TAR or TAR-type models. It seems that the threshold moving average (TMA) model has not attracted too much attention. Maximum likelihood estimation was studied by Gooijer (1998) when the error is normal. However, its large-sample property still remains open. Ling and Tong (2005) considered a quasi-likelihood ratio test for the null linear moving average (MA) model against the TMA model alternative. This test was extended to the heteroskedastic case by Li and Li (2008). Ling, Tong, and Li (2007) obtained the ergodic and invertible conditions for a multiple TMA(1) model. Li, Ling, and Tong (2012) proved that the TMA model is always strictly stationary and ergodic by using a new approach without resorting to Markov chain theory. In practice, there have been many applications in economics and econometrics. For example, Gooijer (1998) used a TMA model to fit the real U.S. gross national product growth rate for the period 1947.I–1982.IV. Ling and Tong (2005) used a TMA(1) model to fit monthly data from January 1971 to December 2000 of the exchange rate of the Japanese yen against the U.S. dollar. These applications also motivate us to further study TMA models in theoretical aspect.

This paper studies the asymptotic theory of LSE in a TMA model. Under some mild conditions, it is shown that the estimator of the threshold is n-consistent and its limiting distribution is related to a two-sided compound Poisson process (CPP), whereas the estimators of other coefficients are strongly consistent and asymptotically normal. This paper also provides a resampling method to tabulate the limiting distribution of the estimated threshold in practice, which is the first successful effort in this direction. Simultaneously, simulation studies are carried out to assess the performance of LSE in finite samples. Unlike the TAR model in Chan (1993), the V-uniform ergodicity of the TMA model is not available in the literature. For the rate of convergence of the estimated threshold, our arguments highly depend on the invertible representation of residuals in Ling and Tong (2005) and are much more complicated than those in TAR models. Because the residual in the TMA model includes infinite threshold indicators, it is hard to imagine the limiting behavior of the profile objective function, and the jump sizes in our CPP are extremely different from those in CPP in Chan (1993). Another key technique in this paper is the weak convergence of a pure jump process. Such weak convergence of a sequence of pure jump processes has been established by Kushner (1984, Sect. 7, pp. 123–126) when the limiting process is a diffusion process. However, when the limiting process is a jump-diffusion process, he only sketches the idea for Markov chains. In the Appendix, we formalize and extend Kushner's idea to non–Markov chain cases when the limiting process is a pure jump process. More importantly, this result is of independent interest by itself and can be applied to many other threshold time series models.

The rest of this paper is organized as follows. Section 2 presents the LSE of the TMA model and states the results. Section 3 considers a resampling method in tabulating the limiting distribution of the estimated threshold when the model parameters are completely unknown. Simulation studies are conducted to assess the performance of LSE in finite samples in Section 4. Some concluding remarks are given in Section 5. Sections 6–9 give proofs of theorems. The Appendix establishes the weak convergence of a pure jump process.

Throughout the paper, some symbols are conventional. The term C is a positive constant, which may be different in different places; $\mathbb{1}(\cdot)$ is the indicator function; $\mathbb{R}^p$ is the euclidean space of dimension p; $\|\cdot\|$ denotes the euclidian norm; $\|\cdot\|_\infty$ is the supremum norm, that is, $\|f\|_\infty = \sup_{x\in\mathbb{R}}|f(x)|$; $o_p(1)$ ($O_p(1)$) denotes a sequence of random numbers converging to zero (bounded) in probability; and $\Longrightarrow$ denotes weak convergence.

2. LEAST SQUARES ESTIMATION AND MAIN RESULTS

A time series $\{y_t, t = 0, \pm 1, ...\}$ is said to be a TMA(1) model if it satisfies the equation

$$y_t = e_t + \left[\phi \mathbb{1}(y_{t-1} \le r) + \psi \mathbb{1}(y_{t-1} > r)\right] e_{t-1}, \tag{2.1}$$

where $\{e_t\}$ is a sequence of independent and identically distributed (i.i.d.) random variables and $r \in \mathbb{R}$ is the threshold parameter. Let $\lambda = (\phi, \psi)'$. Assume that $\lambda \in \Lambda$, a compact subset of $\mathbb{R}^2$, and there exist two finite constants $\underline{r}$ and $\bar{r}$ such that $r \in [\underline{r}, \bar{r}]$ because model (2.1) reduces to a linear MA model when $r = \pm\infty$, which is not of interest in this paper. Here $\theta_0 = (\lambda_0', r_0)'$ is the true parameter of $\theta = (\lambda', r)'$, and it is an interior point in $\Lambda \times [\underline{r}, \bar{r}]$. The parameter space is denoted by $\Theta = \Lambda \times [\underline{r}, \bar{r}]$. Throughout the paper we assume that $\mathbb{E}e_1 = 0$ and $\mathbb{E}e_1^2 < \infty$. Three further assumptions are as follows.

Assumption 2.1. $|\phi| < 1$, $|\psi| < 1$, and Θ is compact.

Assumption 2.2. $\phi_0 \ne \psi_0$.

Assumption 2.3. e_1 has a continuous and strictly positive density $h(x)$ on $\mathbb{R}$ with $\sup_{x\in\mathbb{R}}\{(1+x^4)h(x)\} < \infty$ and $\mathbb{E}e_1^4 < \infty$.

Assumption 2.1 is a sufficient and easy-to-check condition available for the invertibility of model (2.1). When $\phi = \psi$, the invertible region of model (2.1) is the same as that of the MA(1) model. Assumption 2.2 is the identification condition for the threshold r. Assumption 2.3 is a sufficient condition for the strict stationarity and ergodicity of model (2.1); see Li et al. (2012). Under Assumptions 2.1 and 2.3, from Ling and Tong (2005), the residual $e_t(\theta)$ has the following representation:

$$
\begin{aligned}
e_t(\theta) &= y_t - \left[\psi + (\phi - \psi)\,\mathbb{1}\,(y_{t-1} \le r)\right] e_{t-1}(\theta) \\
&= \sum_{j=0}^{\infty} H_{tj}(\theta) y_{t-j},
\end{aligned}
$$

where

$$
H_{tj}(\theta) = \prod_{i=1}^{j} \left[-\psi - (\phi - \psi)\,\mathbb{1}\,(y_{t-i} \le r)\right]
$$

for all $j \ge 0$ with the convention $\prod_{i=1}^{0} \equiv 1$.

Assume that $\{y_1, ..., y_n\}$ is a sample from model (2.1) with sample size n. Given the initial value $Y_0 \equiv \{y_s : s \le 0\}$ or $e_t(\theta) \equiv 0$ for $t \le 0$, the sum of squares errors function $L_n(\theta)$ is defined as

$$
L_n(\theta) = \sum_{t=1}^{n} e_t^2(\theta).
$$

The minimizer $\widehat{\theta}_n = (\widehat{\lambda}_n', \widehat{r}_n)'$ of $L_n(\theta)$ is called the least squares estimator of θ_0, that is,

$$
\widehat{\theta}_n = \arg\min_{\theta \in \Theta} L_n(\theta).
$$

Note that $L_n(\theta)$ is discontinuous in r. The way to get $\widehat{\theta}_n$ is as follows. First, for each fixed $r \in [\underline{r}, \overline{r}]$, we minimize $L_n(\theta)$ and get its minimizer $\widehat{\lambda}_n(r)$. Because $L_n^*(r) \equiv L_n(\theta)|_{\lambda=\widehat{\lambda}_n(r)}$ only takes finite possible values, we then get the minimizer $\widehat{r}_n$ of $L_n^*(r)$ by the enumeration approach. Finally, we can obtain $\widehat{\theta}_n = (\widehat{\lambda}_n'(\widehat{r}_n), \widehat{r}_n)'$. Generally, there exist infinitely many r such that $L_n(\cdot)$ attains its global minimum. One can choose the smallest r as the estimator of r_0. According to the procedure for $\widehat{\theta}_n$, it is not hard to show that $\widehat{\theta}_n$ is the least squares estimator of θ_0.

In practice, however, the initial value Y_0 is not available, and hence we have to replace it by some constants, For example, $Y_0 = x \equiv \{x_1, x_2, ...\}$. Because $\sup_{\theta \in \Theta} \|H_{tj}(\theta)\| = O(\rho^j)$ almost surely (a.s.) for some $\rho \in (0,1)$ by Theorem A.1 in Ling and Tong (2005), we can show that

$$
\sup_{\theta \in \Theta} \left| e_t^2(\theta) - e_t^2(\theta)\,|_{Y_0 = x} \right| = O\left(\rho^t\right) \quad \text{a.s.}
$$

for any given x. Thus, the initial value will not affect the asymptotic properties of $\widehat{\theta}_n$. For simplicity, in what follows, we assume that Y_0 is from model (2.1). In this case, $e_t(\theta_0) = e_t$. Actually, in the numerical optimization of $L_n(\theta)$, we can set the initial values Y_0 equal to the sample mean or directly set $e_t(\theta) \equiv 0$ for $t \le 0$. The following result establishes the strong consistency of $\widehat{\theta}_n$.

THEOREM 2.1. *Suppose that Assumptions 2.1 and 2.2 hold, e_1 has a bounded, continuous and strictly positive density $h(x)$ on $\mathbb{R}$, $\mathbb{E}e_1 = 0$, and $\mathbb{E}e_1^2 < \infty$. Then $\widehat{\theta}_n \to \theta_0$ a.s. as $n \to \infty$.*

Theorem 2.2 shows the convergence rates of $\widehat{r}_n$ and $\widehat{\lambda}_n(r)$.

THEOREM 2.2. *Suppose Assumptions 2.1–2.3 hold. Then*

(i) $n\left(\widehat{r}_n - r_0\right) = O_p(1)$;

(ii) $\sqrt{n} \sup_{|r-r_0|\le B/n} \left\|\widehat{\lambda}_n(r) - \widehat{\lambda}_n(r_0)\right\| = o_p(1)$

for any fixed $B \in (0,\infty)$. *Furthermore, as* $n \to \infty$,

$$\sqrt{n}\left(\widehat{\lambda}_n(\widehat{r}_n) - \lambda_0\right) = \sqrt{n}\left(\widehat{\lambda}_n(r_0) - \lambda_0\right) + o_p(1) \Longrightarrow \mathcal{N}\left(0,\ \sigma^2\Sigma^{-1}\right)$$

where $\sigma^2 = \mathbb{E}e_t^2$ *and* $\Sigma = \mathbb{E}[\left(\partial e_t(\theta_0)/\partial\lambda)(\partial e_t(\theta_0)/\partial\lambda'\right)]$ *with*

$$\frac{\partial e_t(\theta_0)}{\partial\lambda} = -\begin{bmatrix}\mathbb{1}(y_{t-1}\le r_0)\\ \mathbb{1}(y_{t-1}>r_0)\end{bmatrix} e_{t-1} - \left[\psi_0 + (\phi_0-\psi_0)\mathbb{1}(y_{t-1}\le r_0)\right] \times \frac{\partial e_{t-1}(\theta_0)}{\partial\lambda}.$$

Generally, because both σ^2 and Σ are unknown, we can estimate them consistently by

$$\widehat{\sigma}_n^2 = \frac{1}{n}\sum_{t=1}^{n}\tilde{e}_t^2\left(\widehat{\theta}_n\right) \quad \text{and} \quad \widehat{\Sigma}_n = \frac{1}{n}\sum_{t=1}^{n}\frac{\partial\tilde{e}_t\left(\widehat{\theta}_n\right)}{\partial\lambda}\frac{\partial\tilde{e}_t\left(\widehat{\theta}_n\right)}{\partial\lambda'},$$

where $\tilde{e}_t(\theta)$ is defined by

$$\tilde{e}_t(\theta) = y_t - \left[\psi + (\phi-\psi)\mathbb{1}(y_{t-1}\le r)\right]\tilde{e}_{t-1}(\theta)$$

with $\tilde{e}_j(\theta) \equiv 0$ for $j \le 0$. Clearly,

$$\frac{\partial\tilde{e}_t(\theta)}{\partial\lambda} = -\begin{bmatrix}\mathbb{1}(y_{t-1}\le r)\\ \mathbb{1}(y_{t-1}>r)\end{bmatrix}\tilde{e}_{t-1}(\theta) - \left[\psi + (\phi-\psi)\mathbb{1}(y_{t-1}\le r)\right]\frac{\partial\tilde{e}_{t-1}(\theta)}{\partial\lambda}.$$

By some algebraic calculations, we can show that $\widehat{\sigma}_n^2 \to \sigma^2$ and $\widehat{\Sigma}_n \to \Sigma$ in probability.

To study the limiting distribution of $\widehat{r}_n$, we need to consider the following profile sum of squares errors function:

$$\tilde{L}_n(z) = L_n\left(\widehat{\lambda}_n\left(r_0 + \frac{z}{n}\right), r_0 + \frac{z}{n}\right) - L_n\left(\widehat{\lambda}_n(r_0), r_0\right), \qquad z \in \mathbb{R}.$$

We can show that $\tilde{L}_n(z)$ can be approximated in $\mathbb{D}(\mathbb{R})$, the space of all cádlág functions on $\mathbb{R}$ being equipped with the Skorokhod metric, by

$$\wp_n(z) = \mathbb{1}(z<0)\sum_{t=1}^{n}\zeta_{1t}\mathbb{1}\left(r_0 + \frac{z}{n} < y_{t-1} \le r_0\right) + \mathbb{1}(z\ge 0) \times \sum_{t=1}^{n}\zeta_{2t}\mathbb{1}\left(r_0 < y_{t-1} \le r_0 + \frac{z}{n}\right),$$

where

$$\zeta_{1t} = \left\{ \sum_{j=0}^{\infty} H_{t+j,j}^2(\theta_0) \right\} \delta_t^2 + 2 \left\{ \sum_{j=0}^{\infty} e_{t+j} H_{t+j,j}(\theta_0) \right\} \delta_t,$$

$$\zeta_{2t} = \left\{ \sum_{j=0}^{\infty} H_{t+j,j}^2(\theta_0) \right\} \delta_t^2 - 2 \left\{ \sum_{j=0}^{\infty} e_{t+j} H_{t+j,j}(\theta_0) \right\} \delta_t \tag{2.2}$$

with $\delta_t = (\phi_0 - \psi_0) e_{t-1}$.

We define a two-sided CPP $\wp(z)$ as follows:

$$\{\wp(z), z \in \mathbb{R}\} = \{\wp_1(-z)\mathbb{1}(z < 0) + \wp_2(z)\mathbb{1}(z \geq 0),\ z \in \mathbb{R}\}, \tag{2.3}$$

where $\{\wp_1(z), z \geq 0\}$ and $\{\wp_2(z), z \geq 0\}$ are two independent CPPs with $\wp_1(0) = \wp_2(0) = 0$ a.s., with the same jump rate $\pi(r_0) > 0$ (implied by Lemma 6.3), where $\pi(x)$ is the density function of y_1, and with the jump distributions $F_1(\cdot|r_0)$ and $F_2(\cdot|r_0)$, where $F_k(\cdot|r_0)$ is the conditional probability distribution induced by ζ_{k2} given $y_1 = r_0$ for $k = 1, 2$. Because $\int x dF_k(x|r_0) > 0$, $\wp(z)$ goes to $+\infty$ a.s. when $z \to \pm\infty$. Thus, there exists a unique random interval $[M_-, M_+)$ on which the process $\wp(z)$ attains its global minimum a.s. Here, we work with the left continuous version for $\wp_1(z)$ and the right continuous version for $\wp_2(z)$. Now, we have the following theorem.

THEOREM 2.3. *If Assumptions 2.1–2.3 hold, then $n(\widehat{r}_n - r_0) \Longrightarrow M_-$. Furthermore, $n(\widehat{r}_n - r_0)$ is asymptotically independent of $\sqrt{n}(\widehat{\lambda}_n(r_0) - \lambda_0)$, which is always asymptotically normal, regardless of whether r_0 is known or not.*

Compared with the result on the TAR model in Chan (1993), the types of limiting distributions of the estimated thresholds are the same, that is, each of them is the smallest minimizer of a two-sided CPP. However, the biggest essential difference is in the jump distributions of the related CPPs. For the TMA model, an infinite number of threshold indicators are involved in ζ_{kt}'s defined in (2.2), whereas the TAR model has no threshold indicators in jump sizes. On the other hand, ours result and Chan's result are very different from that in Hansen (1997, 2000). Throughout the paper, we consider the case where the threshold effect is fixed and further complete the asymptotic theory on threshold models. When the threshold effect varies with the sample size, Hansen (1997, 2000) has established the corresponding asymptotic theory for TAR models. In this case, it is an interesting and open topic for TMA models, and some further study will be needed in the future.

3. NUMERICAL SIMULATION OF M_-

In this section, we shall provide a resampling method to simulate M_-. Note that ours is the first serious effort in the literature of threshold time series models to estimate the limiting distribution of the threshold estimator. This resampling method contributes to threshold literature and can be used to construct confidence intervals for the threshold parameter in threshold models.

From (2.3), we know that two factors determine the density of M_-, that is, the jump rate $\pi(r_0)$ and the jump distributions $F_1(\cdot|r_0)$ and $F_2(\cdot|r_0)$. We can simulate M_- via simulating the CPP (2.3) on the interval $[-T, T]$ for any given $T > 0$ large enough because the expectations of the jumps are positive. From Algorithm 6.2 in Cont and Tankov (2004, p. 174), we know that the key step is how to sample jump sequences from jump distributions. Because it is impossible to sample jump sequences from $F_k(\cdot|r_0)$ directly, we sample from a consistent estimate of $F_k(\cdot|r_0)$ to replace them. The procedure is as follows.

Given the sample $\mathfrak{X}_n \equiv \{y_1, ..., y_n\}$, we can first use it to estimate θ_0 and $\pi(r_0)$ consistently, denoting the estimators as $\widehat{\theta}_n$ and $\widehat{\pi}(\widehat{r}_n)$, respectively, where $\widehat{\pi}(x)$ is the kernel density estimator of $\pi(x)$, and calculate the residuals $\{\widehat{e}_t : 1 \le t \le n\}$. Based on the residuals, we can construct the estimator $\widehat{h}(x)$ of $h(x)$:

$$\widehat{h}(x) = \frac{1}{nb_n} \sum_{t=1}^{n} K\left(\frac{\widehat{e}_t - x}{b_n}\right),$$

where $K(x) = (\sqrt{2\pi})^{-1}\exp(-x^2/2)$ is the Gaussian kernel and b_n is the bandwidth, which can be selected by

$$b_n = 1.06sn^{-1/5}\left(1 + \frac{35}{48}\widehat{\gamma}_4 + \frac{35}{32}\widehat{\gamma}_3^2 + \frac{385}{1,024}\widehat{\gamma}_4^2\right)^{-1/5},$$

where s, $\widehat{\gamma}_3$, and $\widehat{\gamma}_4$ are the sample standard deviation, skewness, and kurtosis of the residuals $\{\widehat{e}_t : 1 \le t \le n\}$, respectively. See Hjort and Jones (1996). Of course, one can use other kernel functions and bandwidths. When $h(x)$ is uniformly continuous, we have that $\|\widehat{h} - h\|_\infty = o_p(1)$ as $n \to \infty$; see Theorem A in Silverman (1978). We have the following algorithm for sampling from a consistent estimate of $F_1(\cdot|r_0)$.

Algorithm

Step 1. Set $\hat{\mathbf{z}}_i = (y_i, \widehat{e}_i)'$ for $i = 1, ..., n$.

Step 2. For each $i \in \{1, ..., n\}$, sample independently $\{\tilde{e}_t : 2 \le t \le m+2\}$ for some large positive integer m from $\widehat{h}(x)$ given $\mathfrak{X}_n$ and generate $\{\tilde{y}_t : 2 \le t \le m+1\}$ by iterating model (2.1) with the initial value $(\tilde{y}_1, \tilde{e}_1)' = (\widehat{r}_n, \widehat{r}_n - g(\hat{\mathbf{z}}_i, \widehat{\theta}_n))'$ and θ_0 being replaced by $\widehat{\theta}_n$. Then calculate $\tilde{\zeta}_{1,2}^{(m)}$, denoted by $\tilde{\xi}_{1,2}^{(m,i)}$, where

$$\begin{aligned}\tilde{\zeta}_{1,2}^{(m)} = &\left\{\sum_{j=0}^{m}\left(\prod_{i=1}^{j}\left[-\widehat{\psi}_n - \left(\widehat{\phi}_n - \widehat{\psi}_n\right)\mathbb{1}\left(\tilde{y}_{i+1} \le \widehat{r}_n\right)\right]\right)^2\right\}\hat{\delta}_2^2 \\ &+ 2\left\{\sum_{j=0}^{m}\tilde{e}_{j+2}\left(\prod_{i=1}^{j}\left[-\widehat{\psi}_n - \left(\widehat{\phi}_n - \widehat{\psi}_n\right)\mathbb{1}\left(\tilde{y}_{i+1} \le \widehat{r}_n\right)\right]\right)\right\}\hat{\delta}_2\end{aligned}$$

with $\hat{\delta}_2 = (\widehat{\phi}_n - \widehat{\psi}_n)\tilde{e}_1 = (\widehat{\phi}_n - \widehat{\psi}_n)\{\widehat{r}_n - g(\hat{\mathbf{z}}_i, \widehat{\theta}_n)\}$ and

$$g(\hat{\mathbf{z}}_i, \theta) = [\phi \mathbb{1}(y_i \le r) + \psi \mathbb{1}(y_i > r)]\widehat{e}_i. \tag{3.1}$$

Step 3. Calculate all $\widehat{\pi}(\widehat{r}_n|\hat{\mathbf{z}}_i) \equiv \widehat{h}(\widehat{r}_n - g(\hat{\mathbf{z}}_i, \widehat{\theta}_n))$ and sample a U from the conditional discrete density: $\mathbb{P}(U = i|\mathfrak{X}_n) = \widehat{\pi}(\widehat{r}_n|\hat{\mathbf{z}}_i)/\{\sum_{l=1}^n \widehat{\pi}(\widehat{r}_n|\hat{\mathbf{z}}_l)\}$ for $i = 1, ..., n$, conditionally independent of $\{\tilde{e}_t, t \ge 2\}$ given $\mathfrak{X}_n$.

Step 4. Obtain $\tilde{Y}_1 = \tilde{\xi}_{1,2}^{(m,U)}$.

By Lemma 9.1 in Section 9, we can see that $\tilde{Y}_1|_{\mathfrak{X}_n} \sim F_1(x|r_0)$ asymptotically for large enough n and m. By repeating the Algorithm, we can obtain a jump sequence $\{\tilde{Y}_i\}$ that can be regarded as the jump sequence from $F_1(x|r_0)$ asymptotically. Similarly, we can obtain a jump sequence $\{\tilde{Z}_i\}$ from $F_2(x|r_0)$ asymptotically.

For $z \in [-T, T]$, the trajectory of CPP $\widehat{\wp}_n^{(m)}(z)$, an approximation of $\wp(z)$, is given by

$$\widehat{\wp}_n^{(m)}(z) = \mathbb{1}(z < 0)\sum_{i=1}^{N_1} \mathbb{1}(U_i > z)\tilde{Y}_i + \mathbb{1}(z \ge 0)\sum_{j=1}^{N_2} \mathbb{1}(V_j < z)\tilde{Z}_j,$$

where N_1 and N_2 are sampled from the Poisson distribution with the parameter $\widehat{\pi}(\widehat{r}_n)T$ and $\{U_1, ..., U_{N_1}\}$ and $\{V_1, ..., V_{N_2}\}$ are two independent jump time sequences from uniform distributions $U[-T, 0]$ and $U[0, T]$, respectively. Then, we take the smallest minimizer of $\widehat{\wp}_n^{(m)}(z)$ on $[-T, T]$ as an approximate observation $\widehat{M}_n^{(m)}$ of M_-. By repeating the preceding algorithm many times, we can simulate a sequence of observations of M_- and use them to make statistical inference for the threshold. Let $\mathbb{P}_{\mathfrak{X}_n}(\cdot|A) = \mathbb{P}(\cdot|A, \mathfrak{X}_n)$. Then, we have the following theorem.

THEOREM 3.1. *If Assumptions 2.1–2.3 hold, then, in probability,*

$$\lim_{m\to\infty}\lim_{n\to\infty}\left|\mathbb{P}_{\mathfrak{X}_n}\left(\widehat{M}_n^{(m)} \le x\right) - \mathbb{P}(M_- \le x)\right| = 0$$

at each $x \in \mathbb{R}$ *for which* $\mathbb{P}(M_- = x) = 0$.

4. SIMULATIONS

To see whether or not the algorithm in Section 3 does work, we now consider the following TMA(1) model:

$$y_t = e_t + \left[\phi_0 \mathbb{1}(y_{t-1} \le r_0) + \psi_0 \mathbb{1}(y_{t-1} > r_0)\right] e_{t-1}, \tag{4.1}$$

where $(\phi_0, \psi_0, r_0) = (0.8, -0.4, 0.6)$ and $e_t \sim \mathcal{N}(0, 1)$. By the kernel method, we get $\pi(r_0) = 0.2534$, where the sample size is 1,000,000. The Gaussian kernel is used, and the bandwidth is 0.0662. Figure 1a displays one realized path of the two-sided CPP (2.3) under model (4.1). Using 1,000 replications, Figure 1b gives

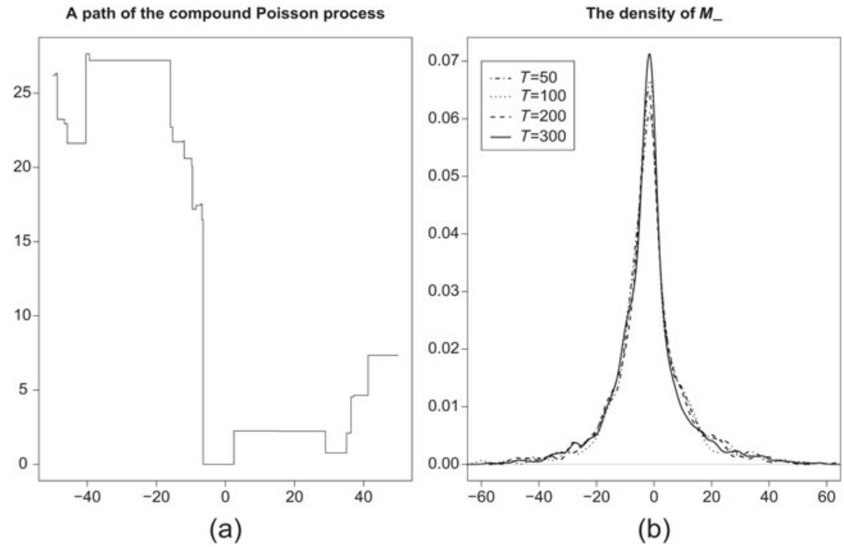

FIGURE 1. (a) One realized path of the two-sided CPP under model (4.1); (b) the density of M_- over different intervals where the CPP is minimized. Here, 1,000 replications are used.

the densities of M_- when $T = 50, 100, 200$, and 300, where $n = 1,000$ and $m = 100$ are used. From the figure, we can see that the densities of M_- are very close to each other for different values of T. According to our simulation experience,

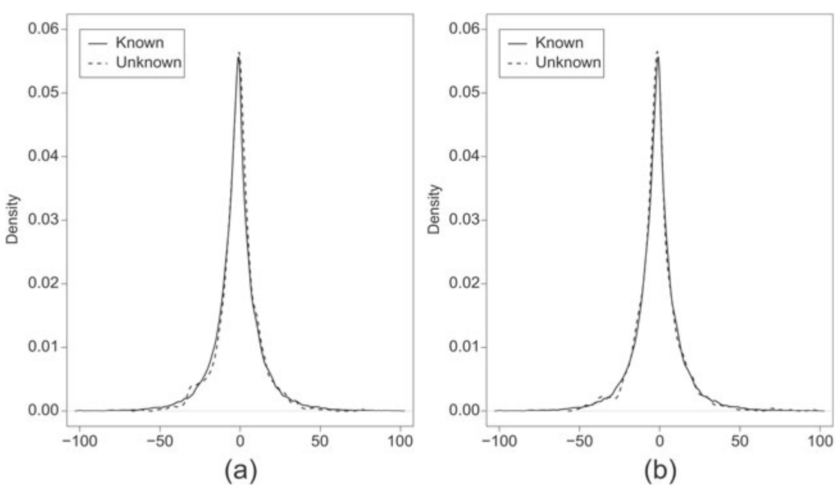

FIGURE 2. The densities of M_- (known) and $\widehat{M}_n^{(m)}$ (unknown), respectively. The sample size is (a) 200 and (b) 400, respectively. Here, $e_t \sim \mathcal{N}(0, 1)$, $T = 300$, and $m = 100$; 10,000 replications are used for M_- and 1,000 replications for $\widehat{M}_n^{(m)}$.

TABLE 1. Empirical quantiles for M_- under model (4.1)

α	0.5%	1%	2.5%	5%	95%	97.5%	99%	99.5%
Q	−63.47	−52.98	−38.25	−27.92	23.06	33.36	55.05	60.72

the larger T is, the more accurate is the estimated density of M_- when $|\phi_0 - \psi_0|$ is small. Unfortunately, there is no theory to support the choice of T in the literature. In practice, we can adopt an attempt for different values of T. For each given one, we can first simulate 100 observations for M_- and plot its density. Based on the support set of the density, we can choose a suitable T. Once T is chosen, we may increase the number of replications to obtain a more precise density of M_-.

Figure 2 displays a more precise density of M_- obtained by 10,000 replications when all parameters are known. When the parameters are unknown, we estimate them by using a given sample $\{y_1, ..., y_n\}$ and then simulate M_-. We can get an approximation $\widehat{M}_n^{(m)}$ of M_-. From the figure, we can see that they are very close even when $n = 200$. This indicates that our resampling method is a useful approach to simulate M_-.

Based on 10,000 replications, Table 1 gives the empirical quantiles of M_- when the significance level $\alpha = 0.005, 0.01, 0.025, 0.05, 0.95, 0.975, 0.99$, and 0.995.

To assess the performance of the least squares estimator of θ_0 in finite samples, we use sample sizes $n = 100$, 200, 400, and 800, each with 1,000 replications for model (4.1). The distribution of e_t is $\mathcal{N}(0,1)$ and student t_5, respectively. In Table 2, we summarize the bias, empirical standard deviation (ESD), and

TABLE 2. Simulation studies for model (4.1) with $\theta_0 = (\phi_0, \psi_0, r_0) = (0.8, -0.4, 0.6)$

		$\mathcal{N}(0,1)$			t_5		
n		ϕ	ψ	r	ϕ	ψ	r
100	Bias	0.0276	0.0005	−0.1410	0.0411	−0.0215	−0.1872
	ESD	0.1085	0.1319	0.2763	0.1254	0.1247	0.3780
	ASD	0.0821	0.1060	0.1710	0.0864	0.1016	0.1791
200	Bias	0.0113	0.0055	−0.0553	0.0180	−0.0088	−0.0556
	ESD	0.0674	0.0844	0.1317	0.0747	0.0826	0.1677
	ASD	0.0581	0.0750	0.0855	0.0611	0.0719	0.0895
400	Bias	0.0035	0.0017	−0.0178	0.0060	−0.0029	−0.0168
	ESD	0.0452	0.0558	0.0426	0.0473	0.0519	0.0634
	ASD	0.0411	0.0530	0.0427	0.0432	0.0508	0.0448
800	Bias	0.0026	−0.0009	−0.0060	0.0042	−0.0032	−0.0045
	ESD	0.0313	0.0385	0.0203	0.0319	0.0360	0.0290
	ASD	0.0290	0.0375	0.0214	0.0306	0.0359	0.0224

TABLE 3. Coverage probabilities when $e_t \sim \mathcal{N}(0,1)$

α	100	200	400	800
0.01	0.951	0.973	0.984	0.989
0.05	0.898	0.925	0.947	0.948
0.10	0.837	0.866	0.897	0.910

asymptotic standard deviation (ASD). Here, the ASDs of $(\widehat{\phi}_n, \widehat{\psi}_n)$ are computed using Σ in Theorem 2.2 and the ASD of $\widehat{r}_n$ is obtained by simulating M_-.

In Table 2, the consistency of the estimators is illustrated by their biases and ESDs. That is, the larger the sample size, the smaller the biases and the closer the ESDs and ASDs on the whole. We also see that the values of the ESDs for $\widehat{r}_n$ are about halved each time when the value of n is doubled. This partially illustrates the n-consistency of the threshold estimator, under which the estimator of the threshold parameter would approach the true parameter much faster than the coefficient parameter estimators.

Table 3 reports the coverage probabilities of r_0 for $n = 100, 200, 400$, and 800, respectively, based on the critical values in Table 1. From the table, we can see that the coverage probability is rather accurate when the sample size n is 400 or above.

Figure 3 shows the empirical distribution and density functions of $n(\widehat{r}_n - r_0)$ when the error is $\mathcal{N}(0,1)$ and the sample size is 800, respectively. From Figure 3,

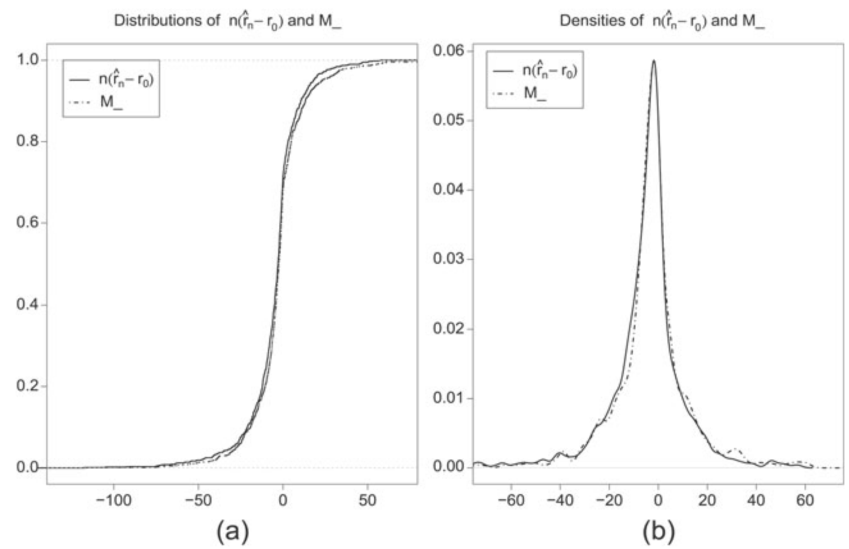

FIGURE 3. The empirical distribution and density functions of both $n(\widehat{r}_n - r_0)$ and M_- for model (4.1) when $e_t \sim \mathcal{N}(0,1)$ and the sample size is 800.

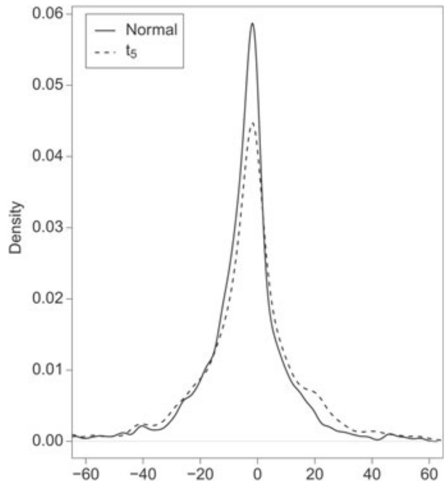

FIGURE 4. The density of $n(\widehat{r}_n - r_0)$ when e_t is $\mathcal{N}(0,1)$ and t_5, respectively. The sample size is 800.

we see that the empirical distribution and density of both $n(\widehat{r}_n - r_0)$ and M_- are very close, which supports Theorem 2.3 empirically. We also see that the density of $n(\widehat{r}_n - r_0)$ is leptokurtic and asymmetric, skewing toward the left-hand side of the origin. Here the skewness is -1.17 and the kurtosis is 10.46. Because of the skewness, confidence intervals of r_0 will not be symmetric about $\widehat{r}_n$, and we must be careful in constructing confidence intervals of the threshold in practice.

From Theorem 2.3, we know that the limiting distribution of $n(\widehat{r}_n - r_0)$ depends on the distribution of the error and is not distribution free. Figure 4 exhibits the density functions of $n(\widehat{r}_n - r_0)$ when the error is $\mathcal{N}(0,1)$ and t_5, respectively. Here, the sample size is 800. From the figure, we can observe the difference for the density functions of $n(\widehat{r}_n - r_0)$ for different types of errors. The density under t_5 is more skewed to the right.

5. CONCLUDING REMARKS

This paper considers a TMA model and has established the asymptotic theory of least squares estimation under some easy-to-verify conditions. We have removed the requirement on the V-uniform ergodicity used in Chan (1993). More importantly, the limiting distribution of the estimated threshold, which is the smallest minimizer of a two-sided compound Poisson process, has been derived. A resampling method is proposed to simulate this limiting distribution, and simulation studies show that it does work well.

For a general TMA(q_1, q_2) model:

$$y_t = \begin{cases} \mu_1 + e_t + \sum\limits_{i=1}^{q_1} \phi_i e_{t-i}, & \text{if } y_{t-d} \le r; \\ \mu_2 + e_t + \sum\limits_{i=1}^{q_2} \psi_i e_{t-i}, & \text{if } y_{t-d} > r; \end{cases}$$

similar results can be easily achieved under some mild conditions when the order (q_1, q_2) is known; see Li, Ling, and Li (2010). In applications, because it is generally unknown, the order (q_1, q_2) can be selected by using the Akaike information criterion (AIC):

$$\text{AIC}(\{q_i\}) = n \log \widehat{\sigma}_n^2 + 2(q_1 + 1) + 2(q_2 + 1),$$

where $\widehat{\sigma}_n^2 = 1/n \sum_{t=1}^n \widehat{e}_t^2$ and $\{\widehat{e}_t\}$ are the residuals when (q_1, q_2) is fixed. A similar AIC is used in Tsay (1998) for TAR models. For more information criteria as model selection tools for threshold models, see Kapetanios (2001).

For model (2.1), the convergence rate is always n if the threshold is identifiable when the threshold effect defined in Hansen (1997, 2000) is fixed. There is no other case like that in Chan and Tsay (1998) for continuous TAR models. As the threshold effect varies with the sample size, this is also an important and interesting topic, which was suggested by an anonymous referee. For this, further study will be needed in the future.

6. PROOF OF THEOREM 2.1

In what follows, $\mathcal{F}_m^n$ denotes a σ-field generated by $\{e_m, ..., e_n\}$ for $m \le n$. Theorem 2.1 can be proved by the approach of Hubers (1967) with the following lemmas. Hence, it is omitted. More details can be found in Li et al. (2010).

LEMMA 6.1. *If Assumption 2.1 holds and* $\mathbb{E}|e_1|^\kappa < \infty$ *for some* $1 \le \kappa \le 4$, *then it follows that*

$$\mathbb{E} \sup_{\theta \in \Theta} |e_t(\theta)|^\kappa < \infty.$$

Proof. From model (2.1), it is not hard to see that $\mathbb{E}|y_t|^\kappa < \infty$. Using the stationarity of $\{y_t\}$ (see Li et al., 2012) and the representation $e_t(\theta) = \sum_{j=0}^\infty H_{tj}(\theta) y_{t-j}$, we can show that the conclusion holds. ∎

LEMMA 6.2. *For any* $\theta \in \Theta$, *define an open neighborhood of* θ *for any* $\eta > 0$ *as*

$$U_\theta(\eta) = \left\{\theta^* = \left(\lambda^*, r^*\right) \in \Theta : \|\lambda^* - \lambda\| < \eta,\ |r^* - r| < \eta\right\}.$$

If Assumption 2.1 holds and $\mathbb{E}e_t^2 < \infty$, *then*

$$\mathbb{E} \sup_{\theta^* \in U_\theta(\eta)} \left| e_t^2\left(\theta^*\right) - e_t^2(\theta) \right| \to 0, \quad \textit{as } \eta \to 0.$$

Proof. The proof is technical. Hence it is omitted. See Li et al. (2010). ∎

LEMMA 6.3.

(i) If the density $h(x)$ of e_1 is continuous and bounded, then the density $\pi(x)$ of y_t is also continuous and bounded.
(ii) If $h(x) > 0$ on $\mathbb{R}$, then $\pi(x) > 0$ and there exist constants $0 < m_0 < M_0 < \infty$ such that $m_0 u \le \mathbb{P}(r < y_0 \le r + u) \le M_0 u$ for fixed $r \in \mathbb{R}$ and any $u \in [0, 1]$.

Proof. From model (2.1), $y_t = e_t + \bar{e}_{t-1}$, where $\bar{e}_{t-1} = \{\psi_0 + (\phi_0 - \psi_0)\mathbb{1}(y_{t-1} \le r_0)\}e_{t-1}$. Let $\bar{G}(\cdot)$ be the distribution of $\bar{e}_{t-1}$. Observe that $\bar{e}_{t-1}$ and e_t are independent, so that

$$\pi(x) = \int_{\mathbb{R}} h(x - z)\, d\bar{G}(z).$$

Using the property of convolution, we can obtain that y_t has a continuous and positive density, which in turn implies that there exist constants $m_0 > 0$ and $M_0 < \infty$ such that

$$m_0 \le \pi(x) \le M_0, \quad \text{as } x \in [r, r+1].$$

Thus, the result holds. ∎

LEMMA 6.4. *If the conditions of Theorem 2.1 hold, then $\mathbb{E}e_t^2(\theta) \ge \sigma^2$ and the equality holds if and only if $\theta = \theta_0$.*

Proof. By Lemma 6.1, $\mathbb{E}e_t^2(\theta) < \infty$. Clearly, $\mathbb{E}e_t^2(\theta_0) = \sigma^2$ because $e_t(\theta_0) = e_t$. A conditional argument yields

$$\begin{aligned}\mathbb{E}\left\{e_t^2(\theta) - e_t^2\right\} &= \mathbb{E}\left\{\left[e_t + \nabla_{t-1}(\theta)\right]^2 - e_t^2\right\} = \mathbb{E}\nabla_{t-1}^2(\theta)\\ &\quad + 2\mathbb{E}\left\{\nabla_{t-1}(\theta)\mathbb{E}\left(e_t | \mathcal{F}_{-\infty}^{t-1}\right)\right\} = \mathbb{E}\nabla_{t-1}^2(\theta) \ge 0,\end{aligned}$$

where $\nabla_{t-1}(\theta) = [\psi_0 + (\phi_0 - \psi_0)\mathbb{1}(y_{t-1} \le r_0)]e_{t-1} - [\psi + (\phi - \psi)\mathbb{1}(y_{t-1} \le r)]e_{t-1}(\theta)$ and it is measurable with respect to $\mathcal{F}_{-\infty}^{t-1}$.

If there exists a θ_* such that $\mathbb{E}\{e_t^2(\theta_*) - e_t^2\} = 0$, then $\nabla_{t-1}(\theta_*) = 0$ a.s. for each t because $\{e_t(\theta_*)\}$ is strictly stationary, and hence $e_t(\theta_*) \equiv e_t + \nabla_{t-1}(\theta_*) = e_t$ a.s. for each t. Thus,

$$\begin{aligned}0 = e_t(\theta_*) - e_t = \big\{&\left[\psi_0 + (\phi_0 - \psi_0)\mathbb{1}(y_{t-1} \le r_0)\right]\\ &- \left[\psi_* + (\phi_* - \psi_*)\mathbb{1}(y_{t-1} \le r_*)\right]\big\}e_{t-1},\end{aligned}$$

which implies that

$$\left[\psi_0 + (\phi_0 - \psi_0)\mathbb{1}(y_{t-1} \le r_0)\right] - \left[\psi_* + (\phi_* - \psi_*)\mathbb{1}(y_{t-1} \le r_*)\right] = 0. \qquad \textbf{(6.1)}$$

Without loss of generality, suppose that $r_* \le r_0$. Then, it follows that from (6.1)

$$(\phi_0 - \psi_*)^2 \mathbb{P}(r_* < y_{t-1} \le r_0) + (\phi_0 - \phi_*)^2 \mathbb{P}(y_{t-1} \le r_*) + (\psi_0 - \psi_*)^2 \\ \times \mathbb{P}(y_{t-1} > r_0) = 0.$$

Note that y_t has a continuous and positive density on $\mathbb{R}$ by Lemma 6.3; by Assumption 2.2 we have $\phi_* = \phi_0$, $\psi_* = \psi_0$, and $r_* = r_0$. Thus, $\theta_* = \theta_0$. ■

7. PROOF OF THEOREM 2.2

All proofs of lemmas in this section are omitted because they are technical. The details can be found in Li et al. (2010). First, we give three lemmas before proving the theorem.

LEMMA 7.1. *If e_t has a continuous density $h(x)$ on $\mathbb{R}$ with $\sup_{x\in\mathbb{R}}\{(1+x^4)h(x)\} < \infty$ and $\mathbb{E}e_1^4 < \infty$, then for fixed $r \in \mathbb{R}$ and $u > 0$*

(i) $\mathbb{E}\{|e_k|^\nu \mathbb{1}(r < y_t \le r+u)\} \le Cu$ *for* $\nu = 1,\ldots,4$*;*
(ii) $\mathbb{E}\{|e_k|^\alpha |e_m|^\beta \mathbb{1}(r < y_t \le r+u)\} \le Cu$ *for* $\alpha, \beta = 1,\ldots,4$ *and* $k \ne m$.

LEMMA 7.2. *If e_1 has a continuous and positive density on $\mathbb{R}$, then there exists a constant $\rho \in (0,1)$ such that for all $t \ge 1$, any $\underline{u}, \bar{u}, \underline{\upsilon}, \bar{\upsilon} \in \bar{\mathbb{R}} \equiv \mathbb{R} \cup \{\pm\infty\}$ and any integer $m \ge 0$,*

$$\Big|\mathbb{E}\left\{g\left(e_j, j \ge 1\right)\mathbb{1}\left(\underline{u} < y_t \le \bar{u}\right)\mathbb{1}\left(\underline{\upsilon} < y_{t+m} \le \bar{\upsilon}\right)\big|\mathcal{F}_{-\infty}^0\right\} \\ -\mathbb{E}\left\{g\left(e_j, j \ge 1\right)\mathbb{1}\left(\underline{u} < y_t \le \bar{u}\right)\mathbb{1}\left(\underline{\upsilon} < y_{t+m} \le \bar{\upsilon}\right)\right\}\Big| \le C\rho^t,$$

where $g(\cdot)$ is a measurable function and satisfies $\mathbb{E}\{g(e_j, j \ge 1)\}^2 < \infty$.

The approach of Chan (1993) highly depended on the V-uniform ergodicity when he studied the convergence rate of the estimated threshold in TAR models. However, for TMA models, the V-uniform ergodicity is not available in the literature. Lemma 7.2 is a counterpart of V-uniform ergodicity, and it plays a key role in the proof of the following lemma.

LEMMA 7.3. *If Assumptions 2.1 and 2.3 hold, then, for any $\varepsilon > 0$ and $\eta > 0$, there exists a constant $B > 0$ such that, for n large enough,*

(i) $\mathbb{P}\left(\sup_{B/n<u\le 1}\left|\frac{\sum_{t=1}^n e_t A_{0t}(u)}{nG(u)}\right| > \eta\right) < \varepsilon$*;*

(ii) $\mathbb{P}\left(\sup_{B/n<u\le 1}\left|\frac{\sum_{t=1}^n [Z_t - \mathbb{E}Z_t]}{nG(u)}\right| < \eta\right) > 1-\varepsilon$*;*

(iii) $\mathbb{P}\left(\sup_{B/n<u\le 1}\left|\frac{\sum_{t=1}^n [W_t(u) - \mathbb{E}W_t(u)]}{nG(u)}\right| > \eta\right) < \varepsilon$*;*

(iv) $\mathbb{P}\left(\sup_{B/n<u\le 1}\left|\frac{\sum_{t=1}^n [K_t(u) - \mathbb{E}K_t(u)]}{nG(u)}\right| > \eta\right) < \varepsilon$*,*

where $G(u) = \mathbb{P}(r_0 < y_0 \le r_0 + u)$, $Z_t = e_{t-1}^2 \mathbb{1}(r_0 < y_{t-1} \le r_0 + u)$,

$$A_{0t}(u) = \sum_{j=0}^{\infty} H_{tj}^0 (r_0 + u) e_{t-j-1} \mathbb{1}\left(r_0 < y_{t-1-j} \le r_0 + u\right),$$

$$W_t(u) = |e_{t-1}| \mathbb{1}(r_0 < y_{t-1} \le r_0 + u) \sum_{j=1}^{\infty} \rho^j |e_{t-1-j}| \times \mathbb{1}\left(r_0 < y_{t-1-j} \le r_0 + u\right),$$

$$K_t(u) = \xi_{\rho t} \left[U_t(0,u) + U_{t-1}(0,u)\right] \quad \textit{with}$$

$$H_{tj}^0(r) = \prod_{i=1}^{j} \left[-\psi_0 - (\phi_0 - \psi_0) \mathbb{1}(y_{t-i} \le r)\right];$$

$$\xi_{0\rho t} = \sum_{i=0}^{\infty} \rho^i |e_{t-i}|; \qquad \xi_{\rho t} = \sum_{i=0}^{\infty} \rho^i |\xi_{0\rho\, t-i}|; \tag{7.1}$$

$$U_t(z_1, z_2) = \sum_{j=0}^{\infty} \rho^{j-1} \xi_{\rho\, t-j} \sum_{i=0}^{j+1} \mathbb{1}(r_0 + z_1 < y_{t-i+1} \le r_0 + z_2)$$

for some $\rho \in (0,1)$.

Proof of Theorem 2.2(i). Because $\widehat{\theta}_n$ is consistent, we restrict the parameter space to an open neighborhood of θ_0. To this end, define $V_\delta = \{\theta \in \Theta : \|\lambda - \lambda_0\| < \delta, |r - r_0| < \delta\}$ for some $0 < \delta < 1$; δ is determined later. Then it suffices to show that there exist constants $B > 0$, $\gamma > 0$, such that for n large enough

$$\mathbb{P}\left(\inf_{\substack{B/n < |r - r_0| \le \delta \\ \theta \in V_\delta}} \frac{L_n(\lambda, r) - L_n(\lambda, r_0)}{nG(|r - r_0|)} > \gamma\right) > 1 - \varepsilon.$$

Here, we only treat the case $r > r_0$. The proof for the case $r < r_0$ is similar.

Write $r = r_0 + u$ for some $0 < u < 1$. Decompose $L_n(\lambda, r) - L_n(\lambda, r_0)$ into two parts, namely,

$$\begin{aligned} L_n(\lambda, r) - L_n(\lambda, r_0) &= \left\{\left[L_n(\lambda, r) - L_n(\lambda, r_0)\right] - [L_n(\lambda_0, r) - L_n(\lambda_0, r_0)]\right\} \\ &\quad + \left[L_n(\lambda_0, r) - L_n(\lambda_0, r_0)\right] \\ &\equiv L_n^{(1)}(\lambda, r) + L_n^{(2)}(r). \end{aligned}$$

We first consider $L_n^{(2)}(r)$. By Theorem A.2 in Ling and Tong (2005), it follows that

$$e_t(\lambda_0, r_0 + u) - e_t = -(\phi_0 - \psi_0)\left[e_{t-1} \mathbb{1}(r_0 < y_{t-1} \le r_0 + u) + A_t(u)\right],$$

where

$$A_t(u) = \sum_{j=1}^{\infty} H_{tj}^0 (r_0 + u) e_{t-j-1} \mathbb{1} \left(r_0 < y_{t-1-j} \le r_0 + u \right),$$

where $H_{tj}^0(r)$ is defined in (7.1). Then, we have

$$\begin{aligned} e_t^2 (\lambda_0, r_0 + u) - e_t^2 \ge (\phi_0 - \psi_0)^2 e_{t-1}^2 \mathbb{1} (r_0 < y_{t-1} \le r_0 + u) - 2 (\phi_0 - \psi_0)^2 W_t(u) \\ - 2 (\phi_0 - \psi_0) e_t A_{0t}(u), \end{aligned} \quad \textbf{(7.2)}$$

where $W_t(u)$ and $A_{0t}(u)$ are defined in Lemma 7.3. A conditional argument and Lemma 6.3 give

$$\sup_{B/n < u \le \delta} \sum_{t=1}^{n} \frac{\mathbb{E} W_t(u)}{n G(u)} \le \sup_{B/n < u \le \delta} \left(\frac{C u^2}{G(u)} \right) = O(\delta). \quad \textbf{(7.3)}$$

On the other hand, there exists a positive constant $m_0 > 0$ such that

$$\lim_{u \downarrow 0} \frac{\mathbb{E} \left\{ e_{t-1}^2 \mathbb{1} (r_0 < y_{t-1} \le r_0 + u) \right\}}{G(u)} \ge m_0. \quad \textbf{(7.4)}$$

By (7.2)–(7.4) and Lemma 7.3(i)–(iii), it follows that

$$\begin{aligned} \inf_{B/n < u \le \delta} \frac{L_n^{(2)}(r_0 + u)}{n G(u)} &\equiv \inf_{B/n < u \le \delta} \left\{ \frac{1}{n G(u)} \sum_{t=1}^{n} \left[e_t^2(\lambda_0, r_0 + u) - e_t^2 \right] \right\} \\ &\ge (\phi_0 - \psi_0)^2 m_0 + O_p(\delta) \end{aligned} \quad \textbf{(7.5)}$$

for sufficiently small $\delta > 0$.

Next, we consider $L_n^{(1)}(\lambda, r)$. Clearly,

$$\frac{1}{n} L_n^{(1)}(\lambda, r) = \frac{1}{n} \sum_{t=1}^{n} \int_0^1 \left[\frac{\partial e_t^2 (\lambda_\upsilon, r)}{\partial \lambda'} - \frac{\partial e_t^2 (\lambda_\upsilon, r_0)}{\partial \lambda'} \right] (\lambda - \lambda_0) \, d\upsilon,$$

where $\lambda_\upsilon = \lambda_0 + \upsilon(\lambda - \lambda_0)$. Using (7.1), we have

$$\left\| \frac{\partial e_t^2(\lambda, r_0 + u)}{\partial \lambda} - \frac{\partial e_t^2(\lambda, r_0)}{\partial \lambda} \right\| \le C \xi_{\rho t} \left\{ U_t(0, u) + U_{t-1}(0, u) \right\} = C K_t(u),$$

where $K_t(u)$ is defined in Lemma 7.3. By Lemma 6.3 and Lemma 7.1, it follows that

$$\sup_{B/n < u \le \delta} \left\{ \sum_{t=1}^{n} \frac{\mathbb{E} K_t(u)}{n G(u)} \right\} \le C. \quad \textbf{(7.6)}$$

From Lemma 7.3(iv) and (7.6), we can obtain

$$\sup_{\substack{B/n<u\le\delta\\ \theta\in V_\delta}} \frac{\left|L_n^{(1)}(\lambda, r_0+u)\right|}{nG(u)} = O_p\left(\sup_{\theta\in V_\delta}\|\lambda-\lambda_0\|\right) = O_p(\delta). \tag{7.7}$$

By (7.5) and (7.7), one can get

$$\begin{aligned}\inf_{\substack{B/n<u\le\delta\\ \theta\in V_\delta}} \frac{L_n(\lambda, r_0+u)-L_n(\lambda, r_0)}{nG(u)} &\ge \inf_{B/n<u\le\delta}\frac{L_n^{(2)}(r_0+u)}{nG(u)}\\ &\quad - \sup_{\substack{B/n<u\le\delta\\ \theta\in V_\delta}}\frac{\left|L_n^{(1)}(\lambda, r_0+u)\right|}{nG(u)}\\ &\ge (\phi_0-\psi_0)^2 m_0 + O_p(\delta).\end{aligned}$$

Let $\gamma = (\phi_0-\psi_0)^2 m_0/2 > 0$. Then, for sufficiently small $\delta > 0$, we have

$$\mathbb{P}\left(\inf_{\substack{B/n<u\le\delta\\ \theta\in V_\delta}} \frac{L_n(\lambda, r_0+u)-L_n(\lambda, r_0)}{nG(u)} > \gamma\right) \ge \mathbb{P}\left(2\gamma + O_p(\delta) > \gamma\right) > 1-\varepsilon.$$

The proof of (i) is complete. ■

For the proof of Theorem 2.2(ii), we need several additional lemmas.

LEMMA 7.4. *For any $\lambda\in\Lambda$, define an open neighborhood of λ for any $\eta > 0$ as $U_\lambda(\eta) = \{\lambda^*\in\Lambda: \|\lambda^*-\lambda\| < \eta\}$. If Assumption 2.1 holds and $\mathbb{E}e_t^2 < \infty$, then*

$$\mathbb{E}\left\{\sup_{r\in[\underline{r},\overline{r}]}\ \sup_{\lambda^*\in U_\lambda(\eta)}\left|e_t^2(\lambda^*, r) - e_t^2(\lambda, r)\right|\right\} \to 0, \quad \textit{as } \eta\to 0.$$

To obtain the limiting distribution of $\widehat{\lambda}_n$, we need to study the uniform convergence of $\widehat{\lambda}_n(r)$ as $r\in[r_0-B/n, r_0+B/n]$ for some $B\in(0,\infty)$.

LEMMA 7.5. *If Assumption 2.1 holds and $\mathbb{E}e_1^2 < \infty$, then for some $0 < B < \infty$*

$$\sup_{|r-r_0|\le B/n}\left\|\widehat{\lambda}_n(r)-\lambda_0\right\| = o_p(1).$$

LEMMA 7.6. *If Assumption 2.1 and 2.3 hold, then, for $\eta > 0$,*

(i) $\mathbb{E}\sup_{\lambda\in\Lambda}\sup_{|r-r_0|<\eta}\left\|\frac{\partial e_t^2(\lambda,r)}{\partial\lambda} - \frac{\partial e_t^2(\lambda,r_0)}{\partial\lambda}\right\| \le C\eta;$

(ii) $\mathbb{E}\sup_{\lambda\in\Lambda}\sup_{|r-r_0|<\eta}\left\|\frac{\partial^2 e_t^2(\lambda,r)}{\partial\lambda\partial\lambda'} - \frac{\partial^2 e_t^2(\lambda,r_0)}{\partial\lambda\partial\lambda'}\right\| \le C\eta.$

LEMMA 7.7. *If Assumptions 2.1 and 2.3 hold, then, for any* $\eta > 0$,

(i) $\mathbb{E} \sup_{\|\lambda-\lambda_0\|<\eta} \sup_{r\in[\underline{r},\overline{r}]} \left\| \frac{\partial e_t^2(\lambda,r)}{\partial\lambda} - \frac{\partial e_t^2(\lambda_0,r)}{\partial\lambda} \right\| \leq C\eta;$

(ii) $\mathbb{E} \sup_{\|\lambda-\lambda_0\|<\eta} \sup_{r\in[\underline{r},\overline{r}]} \left\| \frac{\partial^2 e_t^2(\lambda,r)}{\partial\lambda\partial\lambda'} - \frac{\partial^2 e_t^2(\lambda_0,r)}{\partial\lambda\partial\lambda'} \right\| \leq C\eta.$

LEMMA 7.8. *If Assumptions 2.1 and 2.3 hold, then for any* $0 < B < \infty$

(i) $\sup_{|r-r_0|\leq B/n} \left\| \frac{\partial L_n(\lambda_0,r)}{\partial\lambda} - \frac{\partial L_n(\lambda_0,r_0)}{\partial\lambda} \right\| = O_p(1);$

(ii) $\sup_{\|\lambda-\lambda_0\|<B/\sqrt{n}} \sup_{|r-r_0|\leq B/n} \left\| \frac{\partial^2 L_n(\lambda,r)}{\partial\lambda\partial\lambda'} - \frac{\partial^2 L_n(\lambda_0,r_0)}{\partial\lambda\partial\lambda'} \right\| = O_p\left(n^{1/2}\right).$

Proof of Theorem 2.2(ii). By Taylor's expansion of $\partial L_n(\lambda,r)/\partial\lambda$, we have

$$0 = \frac{1}{n}\frac{\partial L_n\left(\widehat{\lambda}_n,r\right)}{\partial\lambda} = \frac{1}{n}\frac{\partial L_n(\lambda_0,r)}{\partial\lambda} + \frac{1}{n}\frac{\partial^2 L_n\left(\bar{\lambda},r\right)}{\partial\lambda\partial\lambda'}\left[\widehat{\lambda}_n(r)-\lambda_0\right], \tag{7.8}$$

where $\bar{\lambda}$ lies between $\widehat{\lambda}_n(r)$ and λ_0. By the ergodic theorem, it follows that

$$\frac{1}{n}\frac{\partial^2 L_n(\theta_0)}{\partial\lambda\partial\lambda'} \to 2\Sigma, \quad \text{a.s.}$$

as $n\to\infty$. Furthermore, by (7.8) and Lemmas 7.5, 7.7, and 7.8, we have

$$\sup_{|r-r_0|\leq B/n} \left\| \sqrt{n}\left[\widehat{\lambda}_n(r)-\lambda_0\right] + (2\Sigma)^{-1}\frac{1}{\sqrt{n}}\frac{\partial L_n(\theta_0)}{\partial\lambda} \right\| = o_p(1).$$

Hence,

$$\begin{aligned} \sqrt{n}\sup_{|r-r_0|\leq B/n}\left\|\widehat{\lambda}_n(r)-\widehat{\lambda}_n(r_0)\right\| \leq{}& \sup_{|r-r_0|\leq B/n}\left\| \sqrt{n}\left[\widehat{\lambda}_n(r)-\lambda_0\right] + (2\Sigma)^{-1}\frac{1}{\sqrt{n}}\frac{\partial L_n(\theta_0)}{\partial\lambda} \right\| \\ &+ \left\| \sqrt{n}\left[\widehat{\lambda}_n(r_0)-\lambda_0\right] + (2\Sigma)^{-1}\frac{1}{\sqrt{n}}\frac{\partial L_n(\theta_0)}{\partial\lambda} \right\| = o_p(1). \end{aligned}$$

Note that

$$\frac{1}{\sqrt{n}}\frac{\partial L_n(\theta_0)}{\partial\lambda} = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(\frac{\partial e_t^2(\theta_0)}{\partial\phi}, \frac{\partial e_t^2(\theta_0)}{\partial\psi}\right)' = \frac{2}{\sqrt{n}}\sum_{t=1}^{n}\left(\frac{\partial e_t(\theta_0)}{\partial\phi}, \frac{\partial e_t(\theta_0)}{\partial\psi}\right)' e_t$$

and that $\partial L_n(\theta_0)/\partial\lambda$ is a sum of martingale differences in terms of σ-fields $\{\mathcal{F}_{-\infty}^t\}$. From the martingale central limit theorem it follows that

$$\frac{1}{\sqrt{n}}\frac{\partial L_n(\theta_0)}{\partial\lambda} \Longrightarrow \mathcal{N}\left(0,\, 4\sigma^2\Sigma\right) \quad \text{with } \Sigma = \mathbb{E}\left(\frac{\partial e_1(\theta_0)}{\partial\lambda}\frac{\partial e_1(\theta_0)}{\partial\lambda'}\right).$$

Thus, the result holds. The proof is complete. ∎

8. PROOF OF THEOREM 2.3

Before the proof, we first discuss the limiting behavior of the normalized profile sum of squares errors function $\tilde{L}_n(z)$, defined by

$$\tilde{L}_n(z) = L_n\left(\widehat{\lambda}_n\left(r_0 + \frac{z}{n}\right), r_0 + \frac{z}{n}\right) - L_n\left(\widehat{\lambda}_n(r_0), r_0\right), \qquad z \in \mathbb{R}.$$

LEMMA 8.1. *If Assumptions 2.1–2.3 hold, then for any $B \in (0, \infty)$,*

$$\sup_{|z| \le B} \left| \tilde{L}_n(z) - \left[L_n\left(\lambda_0, r_0 + \frac{z}{n}\right) - L_n\left(\lambda_0, r_0\right)\right]\right| = o_p(1).$$

Proof. By Taylor expansion,

$$\begin{aligned}\tilde{L}_n(z) &- \left[L_n\left(\lambda_0, r_0 + \frac{z}{n}\right) - L_n(\lambda_0, r_0)\right] \\ &= \left[\frac{\partial L_n\left(\bar{\lambda}, r_0 + z/n\right)}{\partial \lambda} - \frac{\partial L_n\left(\bar{\lambda}, r_0\right)}{\partial \lambda}\right]\left[\widehat{\lambda}_n\left(r_0 + \frac{z}{n}\right) - \lambda_0\right] \\ &\quad + \frac{\partial L_n\left(\lambda^*, r_0\right)}{\partial \lambda}\left[\widehat{\lambda}_n\left(r_0 + \frac{z}{n}\right) - \widehat{\lambda}_n(r_0)\right],\end{aligned}$$

where $\bar{\lambda}$ lies between λ_0 and $\widehat{\lambda}_n(r_0 + z/n)$ and λ^* lies between $\widehat{\lambda}_n(r_0)$ and $\widehat{\lambda}_n(r_0 + z/n)$. By Lemmas 7.6 and 7.7 and Theorem 2.2, we have

$$\sup_{|z| \le B} \left| \tilde{L}_n(z) - \left[L_n\left(\lambda_0, r_0 + \frac{z}{n}\right) - L_n(\lambda_0, r_0)\right]\right| = o_p(1). \qquad \blacksquare$$

LEMMA 8.2. *If Assumptions 2.1–2.3 hold, then for any $B \in (0, \infty)$,*

$$\mathbb{E} \sup_{|z| \le B} \left| \left[L_n\left(\lambda_0, r_0 + \frac{z}{n}\right) - L_n(\lambda_0, r_0)\right] - \wp_n(z)\right| \to 0 \quad \text{as } n \to \infty.$$

Proof. We only prove the case $r \ge r_0$. The case $r < r_0$ is similar. Let $e_t(z) = e_t(\lambda_0, r_0 + z/n)$. Then, by Theorem A.2 in Ling and Tong (2005), it follows that

$$e_t(z) - e_t = (\psi_0 - \phi_0) \sum_{j=0}^{\infty} H_{tj}(\theta_0) e_{t-1-j} \mathbb{1}\left(r_0 < y_{t-1-j} \le r_0 + \frac{z}{n}\right) + R_t,$$

where $\mathbb{E}|R_t| = O(n^{-2})$. By Theorem A.1 in Ling and Tong (2005), the Hölder inequality, and strict stationarity of $\{y_t\}$, one can get

$$\begin{aligned}[e_t(z) - e_t]^2 = (\psi_0 - \phi_0)^2 &\sum_{j=0}^{\infty} \left[H_{tj}(\theta_0)\right]^2 e_{t-1-j}^2 \mathbb{1}\left(r_0 < y_{t-1-j} \le r_0 + \frac{z}{n}\right) \\ &+ o_p\left(n^{-1}\right).\end{aligned}$$

Then, by exchanging the order of the summation, it follows that

$$L_n\left(\lambda_0, r_0+\frac{z}{n}\right)-L_n(\lambda_0,r_0)=\sum_{t=1}^{n}\left\{[e_t(z)-e_t]^2+2e_t[e_t(z)-e_t]\right\}$$
$$=\sum_{t=1}^{n}\left\{\sum_{j=0}^{\infty}\xi_{t+j,j}\right\}\mathbb{1}\left(r_0<y_{t-1}\le r_0+\frac{z}{n}\right)+o_p(1)$$
$$=\wp_n(z)+o_p(1),$$

where $\xi_{t,j}=(\phi_0-\psi_0)^2H_{tj}^2(\theta_0)e_{t-1-j}^2-2(\phi_0-\psi_0)H_{tj}(\theta_0)e_te_{t-1-j}$ and $o_p(1)$ in this proof is uniform in $z\in[0,B]$ and $\mathbb{E}|o_p(1)|=o(1)$. Thus, the result holds. ■

Proof of Theorem 2.3. In the space $\mathbb{D}(\mathbb{R})$, we use the Skorokhod metric $d(\cdot,\cdot)$, defined as $d(x,y)=\sum_{k=1}^{\infty}2^{-k}\min\{1,d_k(x,y)\}$ for $x,y\in\mathbb{D}(\mathbb{R})$, where $d_k(x,y)$ is the Skorokhod metric in $\mathbb{D}[-k,k]$. See (16.4) and (12.16) in Billingsley (1999, p. 168 and p. 125). Using Lemmas 8.1 and 8.2, we have $d(\tilde{L}_n(z),\wp_n(z))\to 0$ in probability. By Theorem 3.1 in Billingsley (1999, p. 27), it suffices to prove that $\{\wp_n(z),z\in\mathbb{R}\}$ converges weakly as $n\to\infty$. By Theorem 5 in Kushner (1984, p. 32) and Lemma 7.1, it is not hard to show that $\{\wp_n(z),z\in\mathbb{R}\}$ is tight.

Now, we consider a truncated process $\wp_n^{(a)}(z)$ defined by

$$\wp_n^{(a)}(z)=\mathbb{1}\{z<0\}\sum_{t=1}^{n}\zeta_{1t}|_{-a}^{a}\mathbb{1}_t(z,0)+\mathbb{1}\{z\ge 0\}\sum_{t=1}^{n}\zeta_{2t}|_{-a}^{a}\mathbb{1}_t(0,z),$$

where $a>0$, $x|_{-b}^{b}=x\mathbb{1}(|x|\le b)$, and $\mathbb{1}_t(x,y)=\mathbb{1}(r_0+x/n<y_{t-1}\le r_0+y/n)$.

We first show that $\{\wp_n^{(a)}(z),\ z\in\mathbb{R}\}$ converges weakly to a two-sided CPP for each $a>0$. The proof takes two steps: (a) proving the tightness of $\wp_n^{(a)}(z)$; (b) characterizing convergence of finite-dimensional distributions. Note the construction of $\wp_n^{(a)}(z)$. We only deal with the case $z\ge 0$; the opposite case is similar.

(a) *Tightness of* $\wp_n^{(a)}(z)$. By a conditional argument, we have

$$\mathbb{P}(a<y_t\le b,\ c<y_{t-k}\le d)\le C(b-a)(d-c)\quad \text{for } a<b, c<d \quad \text{and} \quad k\ge 1.$$

It is not hard to verify

$$\mathbb{E}\left[\left|\wp_n^{(a)}(z)-\wp_n^{(a)}(z_1)\right|\left|\wp_n^{(a)}(z_2)-\wp_n^{(a)}(z)\right|\right]\le C(z_2-z_1)^2$$

for any $0\le z_1<z<z_2<\infty$. Thus, $\{\wp_n^{(a)}(z),n\ge 1\}$ is tight by Theorem 5 in Kushner (1984, p. 32) again.

(b) *Convergence of finite-dimensional distributions.* For any $0\le z_1\le z_2<z_3\le z_4$ and any real numbers c_1 and c_2, the linear combination of the increments of $\wp_n^{(a)}(z)$ is

$$
\begin{aligned}
S_n &= c_1 \left\{ \wp_n^{(a)}(z_2) - \wp_n^{(a)}(z_1) \right\} + c_2 \left\{ \wp_n^{(a)}(z_4) - \wp_n^{(a)}(z_3) \right\} \\
&= \sum_{t=1}^{n} \zeta_{2t} \Big|_{-a}^{a} \big\{ c_1 \mathbb{1}_t(z_1, z_2) + c_2 \mathbb{1}_t(z_3, z_4) \big\}.
\end{aligned}
$$

Fix $z > 0$ and let $\varepsilon = 1/n$. Consider the following process indexed by ε:

$$
\begin{aligned}
x^{\varepsilon}(t) &= X^{\varepsilon}_{[nt]}, \qquad 0 \le t \le 1, \\
X_0^{\varepsilon} &= 0, \\
X_{k+1}^{\varepsilon} &= X_k^{\varepsilon} + J_{k+1}^{\varepsilon}, \qquad k \ge 1, \\
J_k^{\varepsilon} &= \zeta_{2k} \Big|_{-a}^{a} \big\{ c_1 \mathbb{1}_k(z_1, z_2) + c_2 \mathbb{1}_k(z_3, z_4) \big\},
\end{aligned}
$$

where the symbol $[nt]$ denotes the integral part of nt. Clearly, $x^{\varepsilon}(1) = S_n$. We need to verify Assumptions A.1–A.4 in the Appendix.

First, we have

$$
\begin{aligned}
\mathbb{P}_k^{\varepsilon}\left(J_m^{\varepsilon} \neq 0\right) &= \mathbb{P}\left(J_m^{\varepsilon} \neq 0 | \mathcal{G}_{k-1}\right) \\
&= \mathbb{P}\big(r_0 + z_1 \varepsilon < y_{m-1} \le r_0 + z_2 \varepsilon, \left|\zeta_{2m}\right| \le a | \mathcal{G}_{k-1}\big) \\
&\quad + \mathbb{P}\big(r_0 + z_3 \varepsilon < y_{m-1} \le r_0 + z_4 \varepsilon, \left|\zeta_{2m}\right| \le a \big| \mathcal{G}_{k-1}\big) \\
&= \mathbb{P}\big(|\zeta_{2m}| \le a \big| r_0 + z_1 \varepsilon < y_{m-1} \le r_0 + z_2 \varepsilon, \mathcal{G}_{k-1}\big) \\
&\quad \times \mathbb{P}\big(r_0 + z_1 \varepsilon < y_{m-1} \le r_0 + z_2 \varepsilon \big| \mathcal{G}_{k-1}\big) \\
&\quad + \mathbb{P}\big(|\zeta_{2m}| \le a \big| r_0 + z_3 \varepsilon < y_{m-1} \le r_0 + z_4 \varepsilon, \mathcal{G}_{k-1}\big) \\
&\quad \times \mathbb{P}\big(r_0 + z_3 \varepsilon < y_{m-1} \le r_0 + z_4 \varepsilon \big| \mathcal{G}_{k-1}\big),
\end{aligned}
$$

where $\mathcal{G}_k = \sigma(J_i^{\varepsilon}, i \le k)$. By the strict stationarity of $\{y_t\}$, Lemma 6.3, and Lemma 7.2, it follows that for $j = 1, 3$

$$
\begin{aligned}
&\lim_{\varepsilon \to 0} \lim_{m \to \infty} \varepsilon^{-1} \mathbb{P}\left(r_0 + z_j \varepsilon < y_{m-1} \le r_0 + z_{j+1} \varepsilon | \mathcal{G}_{k-1}\right) = \pi\left(r_0\right)\left(z_{j+1} - z_j\right), \\
&\lim_{\varepsilon \to 0} \lim_{m \to \infty} \mathbb{P}\left(|\zeta_{2m}| \le a | r_0 + z_j \varepsilon < y_{m-1} \le r_0 + z_{j+1} \varepsilon, \mathcal{G}_{k-1}\right) \\
&\qquad = \mathbb{P}\left(|\zeta_{22}| \le a | y_1 = r_0^{+}\right) \equiv \kappa_a.
\end{aligned}
$$

Thus,

$$
\lim_{\varepsilon \to 0} \lim_{m \to \infty} \varepsilon^{-1} \mathbb{P}_k^{\varepsilon}\left(J_m^{\varepsilon} \neq 0\right) = \kappa_a \pi(r_0) \big\{ (z_2 - z_1) + (z_4 - z_3) \big\}. \tag{8.1}
$$

By the stationarity of $\{y_t\}$ again, for any Borel set B, it follows that

$$
\mathbb{Q}^*(B) = \lim_{\varepsilon \to 0} \mathbb{P}\left(J_k^{\varepsilon} \in B | J_k^{\varepsilon} \neq 0\right) = w \mathbb{Q}_1^*(B) + (1 - w) \mathbb{Q}_2^*(B), \tag{8.2}
$$

where $w = (z_2 - z_1)/\{(z_2 - z_1) + (z_4 - z_3)\}$ and $\mathbb{Q}_i^*(B) = \mathbb{P}(c_i \zeta_{22}|_{-a}^{a} \in B | y_1 = r_0^{+}, |\zeta_{22}| \le a)$, $i = 1, 2$. Similarly, by Lemma 7.2, we can verify that, for any $f \in \widehat{C}_0^2$ and a scalar x,

$$\lim_{\varepsilon\to 0}\lim_{m\to\infty}\mathbb{E}_k^{\varepsilon}\left\{f(x+J_m^{\varepsilon})-f(x)\middle|J_m^{\varepsilon}\neq 0\right\}=\lim_{\varepsilon\to 0}\mathbb{E}\left\{f(x+J_k^{\varepsilon})-f(x)\middle|J_k^{\varepsilon}\neq 0\right\}$$
$$=\int\left[f(x+u)-f(x)\right]\mathbb{Q}^*(du). \tag{8.3}$$

By (8.1)–(8.3), Assumptions A.1–A.4 in the Appendix hold. By Theorem A.1, we claim that $x^{\varepsilon}(t)$ converges weakly to a CPP $J(t)$ with jump rate $\kappa_a\pi(r_0)\{(z_2-z_1)+(z_4-\ z_3)\}$ and the jump distribution $\mathbb{Q}^*$. Hence, S_n converges weakly to $J(1)$, a compound Poisson random variable. Now, consider the characteristic function $f_a(t)$ of $J(1)$. Clearly,

$$\begin{aligned} f_a(t)&=\exp\left\{-\kappa_a\pi(r_0)\{(z_2-z_1)+(z_4-z_3)\}\left[1-\int_{\mathbb{R}}e^{itx}\mathbb{Q}^*(dx)\right]\right\}\\ &=\exp\left\{-\kappa_a\pi(r_0)(z_2-z_1)\left[1-\int_{\mathbb{R}}e^{itx}\mathbb{Q}_1^*(dx)\right]\right\}\\ &\quad\times\exp\left\{-\kappa_a\pi(r_0)(z_4-z_3)\left[1-\int_{\mathbb{R}}e^{itx}\mathbb{Q}_2^*(dx)\right]\right\},\end{aligned}$$

which is equal to that of the linear combination $c_1\{\wp^{(a)}(z_2)-\wp^{(a)}(z_1)\}+c_2\{\wp^{(a)}(z_4)-\wp^{(a)}(z_3)\}$ of the independent increments of the CPP

$$\wp^{(a)}(z)=\sum_{i=1}^{N^{(a)}(z)}Y_i,\qquad z\in[0,\infty),$$

where $\{N^{(a)}(z),z\in[0,\infty)\}$ is a Poisson process with jump rate $\kappa_a\pi(r_0)$ and $\{Y_i\}$ is i.i.d. having the distribution $\mathbb{Q}^{(a)}$, where $\mathbb{Q}^{(a)}$ is the induced measure of $\zeta_{22}|_{-a}^{a}$ given $y_1=r_0^+$ and $|\zeta_{22}|\le a$. By the Cramer–Wold device, the finite-dimensional distributions of $\wp_n^{(a)}(z)$ converge weakly to those of the CPP $\{\wp^{(a)}(z)\}$ with jump rate $\kappa_a\pi(r_0)$ and jump distribution $\mathbb{Q}^{(a)}$. Thus, $\wp_n^{(a)}(z)$ converges weakly to $\wp^{(a)}(z)$ as $n\to\infty$ in $\mathbb{D}(\mathbb{R})$ for each $a>0$.

Note that $\kappa_a\to 1$ and $\mathbb{Q}^{(a)}\to F_2(\cdot|r_0)$ as $a\to\infty$. Then, by Theorem 16 in Pollard (1984, p. 134), the CPP $\wp^{(a)}(z)$ converges weakly to a CPP $\wp(z)$ with jump rate $\pi(r_0)$ and jump distribution $F_2(\cdot|r_0)$ as $a\to\infty$. On the other hand, for each $k>0$, we have

$$\lim_{a\to\infty}\lim_{n\to\infty}\mathbb{E}\sup_{|z|\le k}\left|\wp_n(z)-\wp_n^{(a)}(z)\right|=0.$$

Thus, for each $\epsilon>0$,

$$\lim_{a\to\infty}\limsup_{n\to\infty}\mathbb{P}\left(d\left(\wp_n^{(a)}(z),\wp_n(z)\right)>\epsilon\right)=0.$$

By Theorem 3.2 in Billingsley (1999, p. 28), $\wp_n(z)\Longrightarrow\wp(z)$. Thus, $\tilde{L}_n(z)$ converges weakly to $\wp(z)$ in $\mathbb{D}(\mathbb{R})$. The remainder is the same as Theorem 2 in Chan (1993). ■

9. PROOF OF THEOREM 3.1

Before the proof, we first give a technical lemma.

LEMMA 9.1. *If the conditions in Theorem 3.1 hold, then, for each m,*

$$\sup_{x\in\mathbb{R}}\left|\mathbb{P}_{\mathfrak{X}_n}\left(\tilde{Y}_1 \le x\right) - F_1^{(m)}(x|r_0)\right| \to 0 \quad \textit{in probability},$$

where $F_k^{(m)}(x|r_0)$ is the conditional distribution induced by

$$\begin{aligned}\zeta_{k2}^{(m)} = &\left\{\sum_{j=0}^{m}\left(\prod_{i=1}^{j}\left[-\psi_0-(\phi_0-\psi_0)\mathbb{1}\left(y_{i+1}\le r_0\right)\right]\right)^2\right\}\delta_2^2 \\ &-2(-1)^k\left\{\sum_{j=0}^{m} e_{j+2}\left(\prod_{i=1}^{j}\left[-\psi_0-(\phi_0-\psi_0)\mathbb{1}\left(y_{i+1}\le r_0\right)\right]\right)\right\}\delta_2 \qquad \textbf{(9.1)}\end{aligned}$$

with $\delta_2=(\phi_0-\psi_0)e_1$ given $y_1=r_0$.

Proof. Note that

$$\begin{aligned}F_k^{(m)}(x|r_0) &= \int \mathbb{P}\left(\zeta_{k2}^{(m)}\le x\,\middle|\,y_1=r_0, Z_0=\mathbf{z}\right)\frac{\pi(r_0|\mathbf{z})}{\pi(r_0)}Q(d\mathbf{z}) \\ &= \sum_{i=1}^{n}\mathbb{P}\left(\zeta_{k2}^{(m)}\le x\,\middle|\,y_1=r_0,\mathbf{Z}_0=\mathbf{z}_i\right)\frac{\pi(r_0|\mathbf{z}_i)}{\sum_{l=1}^{n}\pi(r_0|\mathbf{z}_l)}+o(1)\quad \text{a.s.,}\end{aligned} \qquad \textbf{(9.2)}$$

uniformly in $x\in\mathbb{R}$ by Theorem 2 in Pollard (1984, p. 8), where $\mathbf{Z}_0=(y_0,e_0)'$, $\mathbf{z}_i\in\mathbb{R}^2$, $Q(\cdot)$ is the distribution of $\mathbf{Z}_0$, and $\pi(r_0|\mathbf{z})$ is the conditional density of y_1 given $\mathbf{Z}_0=\mathbf{z}$. Let

$$\mathbb{P}_{\mathcal{Z}}(Y_1\le x)=\sum_{i=1}^{n}\mathbb{P}\left(\zeta_{12}^{(m)}\le x\,\middle|\,y_1=r_0,\mathbf{Z}_0=\mathbf{z}_i\right)\frac{\pi(r_0|\mathbf{z}_i)}{\sum_{l=1}^{n}\pi(r_0|\mathbf{z}_l)}. \qquad \textbf{(9.3)}$$

By (9.2) and (9.3), it follows that

$$\begin{aligned}\sup_{x\in\mathbb{R}}\left|\mathbb{P}_{\mathfrak{X}_n}\left(\tilde{Y}_1\le x\right)-F_1^{(m)}(x|r_0)\right| &\le \sup_{x\in\mathbb{R}}\left|\mathbb{P}_{\mathfrak{X}_n}\left(\tilde{Y}_1\le x\right)-\mathbb{P}_{\mathcal{Z}}(Y_1\le x)\right| \\ &\quad+\sup_{x\in\mathbb{R}}\left|\mathbb{P}_{\mathcal{Z}}(Y_1\le x)-F_1^{(m)}(x|r_0)\right| \\ &=\sup_{x\in\mathbb{R}}\left|\mathbb{P}_{\mathfrak{X}_n}\left(\tilde{Y}_1\le x\right)-\mathbb{P}_{\mathcal{Z}}(Y_1\le x)\right|+o_p(1).\end{aligned}$$

Thus, it suffices to prove that

$$\sup_{x\in\mathbb{R}}\left|\mathbb{P}_{\mathfrak{X}_n}\left(\tilde{Y}_1\le x\right)-\mathbb{P}_{\mathcal{Z}}(Y_1\le x)\right|=o_p(1).$$

From the Algorithm, it follows that

$$\mathbb{P}_{\mathfrak{X}_n}\left(\tilde{Y}_1 \le x\right) = \sum_{i=1}^{n} \mathbb{P}_{\mathfrak{X}_n}\left(\tilde{\zeta}_{1,2}^{(m)} \le x | \tilde{y}_1 = \widehat{r}_n, \tilde{\mathbf{Z}}_0 = \hat{\mathbf{z}}_i\right) \frac{\widehat{\pi}\left(\widehat{r}_n | \hat{\mathbf{z}}_i\right)}{\sum_{l=1}^{n} \widehat{\pi}\left(\widehat{r}_n | \hat{\mathbf{z}}_l\right)},$$

where $\tilde{\mathbf{Z}}_0 = (\tilde{y}_0, \tilde{e}_0)'$. By a simple calculation, using (9.3), we have

$$\begin{aligned}
&\left|\mathbb{P}_{\mathfrak{X}_n}\left(\tilde{Y}_1 \le x\right) - \mathbb{P}_{\mathcal{Z}}\left(Y_1 \le x\right)\right| \\
&\quad \le \frac{2\sum_{i=1}^{n}\left|\widehat{\pi}\left(\widehat{r}_n|\hat{\mathbf{z}}_i\right) - \pi\left(r_0|\mathbf{z}_i\right)\right|}{\sum_{l=1}^{n}\pi\left(r_0|\mathbf{z}_l\right)} \\
&\qquad + \frac{\left\{\max\limits_{1\le i\le n}\widehat{\pi}\left(\widehat{r}_n|\hat{\mathbf{z}}_i\right)\right\}\sum_{i=1}^{n}\left|\mathbb{P}_{\mathfrak{X}_n}\left(\tilde{\zeta}_{1,2}^{(m)} \le x\middle|\tilde{y}_1 = \widehat{r}_n, \tilde{\mathbf{Z}}_0 = \hat{\mathbf{z}}_i\right) - \mathbb{P}\left(\zeta_{12}^{(m)} \le x\middle|y_1 = r_0, \mathbf{Z}_0 = \mathbf{z}_i\right)\right|}{\sum_{l=1}^{n}\widehat{\pi}\left(\widehat{r}_n|\hat{\mathbf{z}}_l\right)}.
\end{aligned}$$

Define the residual $\widehat{e}_t$ by the equation

$$\widehat{e}_t = y_t - \left[\widehat{\phi}_n \mathbb{1}\left(y_{t-1} \le \widehat{r}_n\right) + \widehat{\psi}_n \mathbb{1}\left(y_{t-1} > \widehat{r}_n\right)\right]\widehat{e}_{t-1}$$

with $\widehat{e}_j \equiv 0$ for $j \le 0$. From Theorems 2.1 and 2.2 and Lemma 7.1, some calculations yield that, for $[\sqrt{n}] \le t \le n$, where $[\sqrt{n}]$ denotes the integral part of $\sqrt{n}$,

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left|\widehat{e}_t - e_t\right| = O_p(1) \quad \text{and} \quad \left|\widehat{e}_t - e_t\right| = o_p(1). \tag{9.4}$$

Note that

$$\begin{aligned}
\left|g\left(\hat{\mathbf{z}}_t, \widehat{\theta}_n\right) - g(\mathbf{z}_t, \theta_0)\right| \le &\left(\left|\widehat{\phi}_n - \phi_0\right| + \left|\widehat{\psi}_n - \psi_0\right|\right)|e_t| + \left(\left|\widehat{\phi}_n\right| + \left|\widehat{\psi}_n\right|\right)\left|\widehat{e}_t - e_t\right| \\
&+ \left(\left|\widehat{\phi}_n\right| + \left|\widehat{\psi}_n\right|\right)\mathbb{1}\left(|y_t - r_0| \le \left|\widehat{r}_n - r_0\right|\right)|e_t|.
\end{aligned}$$

We have that for $[\sqrt{n}] \le t \le n$

$$\left|g\left(\hat{\mathbf{z}}_t, \widehat{\theta}_n\right) - g(\mathbf{z}_t, \theta_0)\right| = o_p(1). \tag{9.5}$$

Let

$$\tilde{h}(x) = \frac{1}{nb_n}\sum_{t=1}^{n} K\left(\frac{e_t - x}{b_n}\right).$$

Then, by the mean value theorem,

$$\left\|\widehat{h} - \tilde{h}\right\|_\infty \le \frac{1}{\sqrt{n}\, b_n^2}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left|\widehat{e}_t - e_t\right| = o_p(1)$$

by (9.4) and $b_n \sim n^{-1/5}$. By Theorem A in Silverman (1978), $\|\tilde{h}-h\|_\infty = o_p(1)$. Thus,

$$\left\|\widehat{h}-h\right\|_\infty \le \left\|\widehat{h}-\tilde{h}\right\|_\infty + \left\|\tilde{h}-h\right\|_\infty = o_p(1).$$

Then, it follows that

$$\begin{aligned}\max_{1\le i\le n} \widehat{\pi}\left(\widehat{r}_n|\hat{\mathbf{z}}_i\right) &\le \max_{1\le i\le n}\left|\widehat{h}\left(\widehat{r}_n - g\left(\hat{\mathbf{z}}_i,\widehat{\theta}_n\right)\right) - h\left(\widehat{r}_n - g\left(\hat{\mathbf{z}}_i,\widehat{\theta}_n\right)\right)\right| + \|h\|_\infty \\ &\le \left\|\widehat{h}-h\right\|_\infty + \|h\|_\infty = \|h\|_\infty + o_p(1)\end{aligned}$$

and

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\left|\widehat{\pi}\left(\widehat{r}_n|\hat{\mathbf{z}}_i\right) - \pi\left(r_0|\mathbf{z}_i\right)\right| &\le \left\|\widehat{h}-h\right\|_\infty + \frac{1}{n}\sum_{i=1}^{n}\left|h\left(\widehat{r}_n - g\left(\hat{\mathbf{z}}_i,\widehat{\theta}_n\right)\right) - h\left(r_0 - g(\mathbf{z}_i,\theta_0)\right)\right| \\ &= o_p(1)\end{aligned}$$

because $\mathbb{E}\left|h(\widehat{r}_n - g(\hat{\mathbf{z}}_i,\widehat{\theta}_n)) - h(r_0 - g(\mathbf{z}_i,\theta_0))\right| = o(1)$ as for $[\sqrt{n}] \le i \le n$ by Theorem 2.1, (9.5), and the uniform continuity of $h(x)$ (implied by Assumption 2.3). Furthermore,

$$\frac{1}{n}\sum_{l=1}^{n}\widehat{\pi}\left(\widehat{r}_n|\hat{\mathbf{z}}_l\right) = \frac{1}{n}\sum_{i=1}^{n}\pi\left(r_0|\mathbf{z}_i\right) + \frac{1}{n}\sum_{i=1}^{n}\left[\widehat{\pi}\left(\widehat{r}_n|\hat{\mathbf{z}}_i\right) - \pi\left(r_0|\mathbf{z}_i\right)\right] = \pi\left(r_0\right) + o_p(1)$$

by the law of large numbers. Therefore,

$$\begin{aligned}\sup_{x\in\mathbb{R}}\left|\mathbb{P}_{\mathfrak{X}_n}\left(\tilde{Y}_1 \le x\right) - \mathbb{P}_{\mathcal{Z}}(Y_1 \le x)\right| \le O_p(1)\Bigg\{\frac{1}{n}\sum_{i=1}^{n}\sup_{x\in\mathbb{R}}\Big|&\mathbb{P}_{\mathfrak{X}_n}\left(\tilde{\zeta}_{1,2}^{(m)} \le x\,\middle|\,\tilde{y}_1 = \widehat{r}_n, \tilde{\mathbf{Z}}_0 = \hat{\mathbf{z}}_i\right) \\ &- \mathbb{P}\left(\zeta_{12}^{(m)} \le x\,\middle|\,y_1 = r_0, \mathbf{Z}_0 = \mathbf{z}_i\right)\Big|\Bigg\} + o_p(1),\end{aligned}$$

where $O_p(1)$ and $o_p(1)$ are uniform in x. Because the difference of two conditional probabilities is bounded by 1, it suffices to prove that for each $[\sqrt{n}] \le i \le n$

$$\sup_{x\in\mathbb{R}}\left|\mathbb{P}_{\mathfrak{X}_n}\left(\tilde{\zeta}_{1,2}^{(m)} \le x\,\middle|\,\tilde{y}_1 = \widehat{r}_n, \tilde{\mathbf{Z}}_0 = \hat{\mathbf{z}}_i\right) - \mathbb{P}\left(\zeta_{12}^{(m)} \le x\,\middle|\,y_1 = r_0, \mathbf{Z}_0 = \mathbf{z}_i\right)\right| = o_p(1). \tag{9.6}$$

Next, we shall prove (9.6). Let $\widetilde{H}_{[k]}(\cdot)$ be the conditional distribution of $\widetilde{\mathbf{Y}}_k \equiv (\tilde{y}_k, ..., \tilde{y}_2)'$ given $\tilde{y}_1 = \widehat{r}_n$, $\tilde{\mathbf{Z}}_0 = \hat{\mathbf{z}}_i$, and $\mathfrak{X}_n$, and let $H_{[k]}(\cdot)$ be the conditional distribution of $\mathbf{Y}_k \equiv (y_k, ..., y_2)'$ given $y_1 = r_0$ and $\mathbf{Z}_0 = \mathbf{z}_i$. By induction over k $(\le m+1)$, we first show that $\|\widetilde{H}_{[m+1]} - H_{[m+1]}\|_\infty = o_p(1)$ as $n \to \infty$.

When $k=2$, it follows that

$$
\begin{aligned}
&\left\|\widetilde{H}_{[2]}-H_{[2]}\right\|_{\infty}\\
&\quad=\sup_{x\in\mathbb{R}}\left|\mathbb{P}_{\mathfrak{X}_n}\left(\tilde{e}_2+g\left(\hat{\mathbf{v}}_i,\widehat{\theta}_n\right)\le x|\tilde{y}_1=\widehat{r}_n,\tilde{\mathbf{Z}}_0=\hat{\mathbf{z}}_i\right)\right.\\
&\qquad\left.-\mathbb{P}\left(e_2+g(\mathbf{v}_i,\theta_0)\le x|y_1=r_0,\mathbf{Z}_0=\mathbf{z}_i\right)\right|\\
&\quad=\sup_{x\in\mathbb{R}}\left|\int_{-\infty}^{x-g\left(\hat{\mathbf{v}}_i,\widehat{\theta}_n\right)}\widehat{h}(u)du-\int_{-\infty}^{x-g(\mathbf{v}_i,\theta_0)}h(u)du\right|\\
&\quad\le\int_{\mathbb{R}}\left|\widehat{h}(u)-h(u)\right|du+\|h\|_{\infty}\left|g\left(\hat{\mathbf{v}}_i,\widehat{\theta}_n\right)-g(\mathbf{v}_i,\theta_0)\right|,
\end{aligned}
$$

where $\hat{\mathbf{v}}_i=(\widehat{r}_n,\widehat{r}_n-g(\hat{\mathbf{z}}_i,\widehat{\theta}_n))'$, $\mathbf{v}_i=(r_0,r_0-g(\mathbf{z}_i,\theta_0))'$, and $g(\cdot,\cdot)$ is defined in (3.1). By the dominated convergence theorem and $\|\widehat{h}-h\|_{\infty}=o_p(1)$, we have $\int_{\mathbb{R}}|\widehat{h}(u)-h(u)|du=o_p(1)$. On the other hand, we have

$$
\begin{aligned}
\left|g\left(\hat{\mathbf{v}}_i,\widehat{\theta}_n\right)-g(\mathbf{v}_i,\theta_0)\right|\le\left(\left|\widehat{\phi}_n-\phi_0\right|+\left|\phi_0\right|\right)&\left\{\left|\widehat{r}_n-r_0\right|+\left|g\left(\hat{\mathbf{z}}_i,\widehat{\theta}_n\right)-g(\mathbf{z}_i,\theta_0)\right|\right\}\\
&+\left|\widehat{\phi}_n-\phi_0\right|\left|r_0-g(\mathbf{z}_i,\theta_0)\right|.
\end{aligned}
$$

By Theorem 2.1 and (9.5), we have $|g(\hat{\mathbf{v}}_i,\widehat{\theta}_n)-g(\mathbf{v}_i,\theta_0)|=o_p(1)$. Thus, $\|\widetilde{H}_{[2]}-H_{[2]}\|_{\infty}=o_p(1)$.

Suppose that $\|\widetilde{H}_{[k]}-H_{[k]}\|_{\infty}=o_p(1)$ for some $k>2$. Consider the case $k+1$. Let $\widetilde{\mathbf{Z}}_t=(\tilde{y}_t,\tilde{e}_t)'$. From the structure of g in (3.1), there exists a piecewise continuous function $f(\cdot,\cdot)$ with at most 2^k segments such that $g(\widetilde{\mathbf{Z}}_k,\widehat{\theta}_n)=f(\widetilde{\mathbf{Y}}_k,\widehat{\theta}_n)$ given $\tilde{y}_1=\widehat{r}_n$, $\tilde{\mathbf{Z}}_0=\hat{\mathbf{z}}_i$, and $\mathfrak{X}_n$, and $g(\mathbf{Z}_k,\theta_0)=f(\mathbf{Y}_k,\theta_0)$ given $y_1=r_0$ and $\mathbf{Z}_0=\mathbf{z}_i$. Then for $A=\prod_{i=1}^{k-1}(-\infty,x_i]$

$$
\begin{aligned}
&\left|\mathbb{P}_{\mathfrak{X}_n}\left(\tilde{y}_{k+1}\le x,\widetilde{\mathbf{Y}}_k\in A\middle|\tilde{y}_1=\widehat{r}_n,\tilde{\mathbf{Z}}_0=\hat{\mathbf{z}}_i\right)-\mathbb{P}\left(y_{k+1}\le x,\mathbf{Y}_k\in A\middle|y_1=r_0,\mathbf{Z}_0=\mathbf{z}_i\right)\right|\\
&\quad=\left|\int_{\mathbb{R}}\mathbb{P}_{\mathfrak{X}_n}\left(f\left(\widetilde{\mathbf{Y}}_k,\widehat{\theta}_n\right)\le x-u,\widetilde{\mathbf{Y}}_k\in A\middle|\tilde{y}_1=\widehat{r}_n,\tilde{\mathbf{Z}}_0=\hat{\mathbf{z}}_i\right)\widehat{h}(u)du\right.\\
&\qquad\left.-\int_{\mathbb{R}}\mathbb{P}\left(f(\mathbf{Y}_k,\theta_0)\le x-u,\mathbf{Y}_k\in A\middle|y_1=r_0,\mathbf{Z}_0=\mathbf{z}_i\right)h(u)du\right|\\
&\quad\le\int_{\mathbb{R}}\left|\widehat{h}(u)-h(u)\right|du+\int_{\mathbb{R}}J_{nk}(u)h(u)du+\int_{\mathbb{R}}I_{nk}(u)h(u)du,
\end{aligned}
$$

where

$$
\begin{aligned}
J_{nk}(u)=&\left|\mathbb{P}_{\mathfrak{X}_n}\left(f\left(\widetilde{\mathbf{Y}}_k,\widehat{\theta}_n\right)\le x-u,\widetilde{\mathbf{Y}}_k\in A\middle|\tilde{y}_1=\widehat{r}_n,\tilde{\mathbf{Z}}_0=\hat{\mathbf{z}}_i\right)\right.\\
&\left.-\mathbb{P}_{\mathfrak{X}_n}\left(f\left(\widetilde{\mathbf{Y}}_k,\theta_0\right)\le x-u,\widetilde{\mathbf{Y}}_k\in A\middle|\tilde{y}_1=\widehat{r}_n,\tilde{\mathbf{Z}}_0=\hat{\mathbf{z}}_i\right)\right|,\\
I_{nk}(u)=&\left|\mathbb{P}_{\mathfrak{X}_n}\left(f\left(\widetilde{\mathbf{Y}}_k,\theta_0\right)\le x-u,\widetilde{\mathbf{Y}}_k\in A\middle|\tilde{y}_1=\widehat{r}_n,\tilde{\mathbf{Z}}_0=\hat{\mathbf{z}}_i\right)\right.\\
&\left.-\mathbb{P}\left(f(\mathbf{Y}_k,\theta_0)\le x-u,\mathbf{Y}_k\in A\middle|y_1=r_0,\mathbf{Z}_0=\mathbf{z}_i\right)\right|.
\end{aligned}
$$

Noting that

$$\begin{aligned}
&\mathbb{E}_{\mathfrak{X}_n}\left\{\left|f\left(\widetilde{\mathbf{Y}}_k,\widehat{\theta}_n\right)-f\left(\widetilde{\mathbf{Y}}_k,\theta_0\right)\right|\,\middle|\,\tilde{y}_1=\widehat{r}_n,\tilde{\mathbf{Z}}_0=\hat{\mathbf{z}}_i\right\}\\
&\quad\le\left(\left|\widehat{\phi}_n-\phi_0\right|+\left|\widehat{\psi}_n-\psi_0\right|\right)\int|x|\widehat{h}(x)dx\\
&\qquad+\left(2\left\|\widehat{h}\right\|_\infty\right)^{1/2}\left(\left|\widehat{\phi}_n\right|+\left|\widehat{\psi}_n\right|\right)\left|\widehat{r}_n-r_0\right|^{1/2}\left(\int|x|^2\widehat{h}(x)dx\right)^{1/2}
\end{aligned}$$

by the Hölder inequality and a conditional argument. Because

$$\left(\int|x|\widehat{h}(x)dx\right)^2\le\int|x|^2\widehat{h}(x)dx=\frac{1}{n}\sum_{t=1}^n\widehat{e}_t^2+b_n^2\to\sigma^2,$$

by the expression of $\widehat{h}(x)$, we have

$$\mathbb{E}_{\mathfrak{X}_n}\left\{\left|f\left(\widetilde{\mathbf{Y}}_k,\widehat{\theta}_n\right)-f(\widetilde{\mathbf{Y}}_k,\theta_0)\right|\,\middle|\,\tilde{y}_1=\widehat{r}_n,\tilde{\mathbf{Z}}_0=\hat{\mathbf{z}}_i\right\}=o_p(1)$$

by Theorem 2.1. That is, $f(\widetilde{\mathbf{Y}}_k,\widehat{\theta}_n)-f(\widetilde{\mathbf{Y}}_k,\theta_0)=o_p(1)$ conditionally on $\tilde{y}_1=\widehat{r}_n$, $\tilde{\mathbf{Z}}_0=\hat{\mathbf{z}}_i$, and $\mathfrak{X}_n$, in probability. Thus, a.s. for all u, $J_{nk}(u)=o_p(1)$. Thanks to $J_{nk}(u)\le 1$, we have that, a.s. for all u, $\mathbb{E}J_{nk}(u)=o(1)$. Then, $\mathbb{E}\int_{\mathbb{R}}J_{nk}(u)h(u)du=\int_{\mathbb{R}}(\mathbb{E}J_{nk}(u))h(u)du=o(1)$ by Fubini's theorem, the dominated convergence theorem, and the continuity of $h(x)$. Thus, $\int_{\mathbb{R}}J_{nk}(u)h(u)du=o_p(1)$.

Because $\|\widetilde{H}_{[k]}-H_{[k]}\|_\infty=o_p(1)$ and $f(\cdot,\theta_0)$ is piecewise continuous, by the continuous mapping theorem, we have that, a.s. for all u, in probability $I_{nk}(u)\to 0$ as $n\to\infty$. By Fubini's theorem and the dominated convergence theorem again, we have $\mathbb{E}\int_{\mathbb{R}}I_{nk}(u)h(u)du=o(1)$. Hence, $|\widetilde{H}_{[k+1]}(\mathbf{x})-H_{[k+1]}(\mathbf{x})|=o_p(1)$ for each $\mathbf{x}\in\mathbb{R}^k$. Because $H_{[k+1]}(\mathbf{y})$ is continuous uniformly in $\mathbf{z}_i$, we have $\|\widetilde{H}_{[k+1]}-H_{[k+1]}\|_\infty=o_p(1)$. Thus, $\|\widetilde{H}_{[m+1]}-H_{[m+1]}\|_\infty=o_p(1)$. From its structure, $\zeta_{k2}^{(m)}$ in (9.1) is a piecewise continuous function of e_{m+2} and $\mathbf{Y}_{m+1}$. (Note: e_j is a piecewise function of $\mathbf{Y}_j$ for $2\le j\le m+1$.) As a result of the independence between e_{m+2} and $\mathbf{Y}_{m+1}$ and the continuity of $h(x)$, $\mathbb{P}(\zeta_{12}^{(m)}\le x|y_1=r_0,\mathbf{Z}_0=\mathbf{z}_i)$ is continuous uniformly in $\mathbf{z}_i$. By the continuous mapping theorem and the continuity of $\mathbb{P}(\zeta_{12}^{(m)}\le x|y_1=r_0,\mathbf{Z}_0=\mathbf{z}_i)$, (9.6) holds.

Therefore, $\sup_{x\in\mathbb{R}}|\mathbb{P}_{\mathfrak{X}_n}(\tilde{Y}_1\le x)-\mathbb{P}_{\mathcal{Z}}(Y_1\le x)|=o_p(1)$. The proof is complete. ∎

Proof of Theorem 3.1. The proof is argued through subsequences. Let $(\Omega,\mathfrak{F},\mathbb{P})$ be the basic probability space. Denote $\pi_n=\widehat{\pi}(\widehat{r}_n)$, $\widehat{F}_{1n}^{(m)}(x)=\mathbb{P}_{\mathfrak{X}_n}(\tilde{Y}_1\le x)$, and $\widehat{F}_{2n}^{(m)}(x)=\mathbb{P}_{\mathfrak{X}_n}(\tilde{Z}_1\le x)$, all of which depend on $\omega\in\Omega$ because $\mathfrak{X}_n$ is defined on Ω. We define a two-sided CPP by $\widehat{\wp}_n^{(m)}(z)$ that is determined by the jump rate π_n and jump distributions $\widehat{F}_{1n}^{(m)}(x)$ and $\widehat{F}_{2n}^{(m)}(x)$. More specifically,

$$\widehat{\wp}_n^{(m)}(z) = I(z<0)\sum_{k=1}^{\widehat{N}_{1n}(-z)} \tilde{Y}_k + I(z \geq 0)\sum_{k=1}^{\widehat{N}_{2n}(z)} \tilde{Z}_k$$
$$\equiv I(z<0)\widehat{\wp}_{1n}(-z) + I(z>0)\widehat{\wp}_{2n}(z),$$

where both $\{\widehat{N}_{1n}(z) : z \geq 0\}$ and $\{\widehat{N}_{2n}(z) : z \geq 0\}$ are conditional independent CPPs given $\mathfrak{X}_n$ and have the same jump rate π_n, $\{\tilde{Y}_k\}$ are i.i.d. from $\widehat{F}_{1n}^{(m)}(x)$ and $\{\tilde{Z}_k\}$ i.i.d. from $\widehat{F}_{2n}^{(m)}(x)$, and $\{\tilde{Y}_k\}$ and $\{\tilde{Z}_k\}$ are mutually conditional independent given $\mathfrak{X}_n$. Suppose that $\widehat{\wp}_n^{(m)}(z)$ is defined on the probability space $(\widetilde{\Omega}, \widetilde{\mathfrak{F}}, \widetilde{\mathbb{P}})$. To stress the dependence of $\widehat{\wp}_n^{(m)}$ on $\tilde{\omega}$, z, and ω (or on ω), we write it as $\widehat{\wp}_n^{(m)}(\tilde{\omega}, z; \omega)$ (or $\widehat{\wp}_n^{(m)}(\omega)$). Because π_n is a weakly consistent estimator of $\pi(r_0)$, it follows that $\left(\pi_n, \widehat{F}_{1n}^{(m)}(x), \widehat{F}_{2n}^{(m)}(x)\right) \to (\pi(r_0), F_1^{(m)}(x|r_0), F_2^{(m)}(x|r_0))$ in probability by Lemma 9.1, uniformly in $x \in \mathbb{R}$. Thus, for any subsequence $(\pi_{n_k}, \widehat{F}_{1n_k}^{(m)}(x), \widehat{F}_{2n_k}^{(m)}(x))$, there exists a further subsequence $(\pi_{n_{k_i}}, \widehat{F}_{1n_{k_i}}^{(m)}(x), \widehat{F}_{2n_{k_i}}^{(m)}(x))$ such that

$$\left(\pi_{n_{k_i}}, \widehat{F}_{1n_{k_i}}^{(m)}(x), \widehat{F}_{2n_{k_i}}^{(m)}(x)\right) \to \left(\pi(r_0), F_1^{(m)}(x|r_0), F_2^{(m)}(x|r_0)\right), \quad \mathbb{P}\text{-a.s.},$$

uniformly in $x \in \mathbb{R}$. Then, there exists a subset $\mathcal{A} \subset \Omega$ with $\mathbb{P}(\mathcal{A}^c) = 0$ such that for each fixed $\omega \in \mathcal{A}$

$$\left(\pi_{n_{k_i}}(\omega), \widehat{F}_{1n_{k_i}}^{(m)}(x)(\omega), \widehat{F}_{2n_{k_i}}^{(m)}(x)(\omega)\right) \to \left(\pi(r_0), F_1^{(m)}(x|r_0), F_2^{(m)}(x|r_0)\right), \quad \textbf{(9.7)}$$

uniformly in $x \in \mathbb{R}$. Consider the process $\widehat{\wp}_{n_{k_i}}^{(m)}(\tilde{\omega}, z; \omega)$. For each fixed $\omega \in \mathcal{A}$, in what follows, we will show that $\widehat{\wp}_{n_{k_i}}^{(m)}(\tilde{\omega}, z; \omega)$ converges weakly by Theorem 16 in Pollard (1984, p. 134). To this end, we need to verify the following two conditions.

(a) *Aldous's condition*. Because every CPP is a Lévy process, by the strong Markov property of Lévy processes (see Cont and Tankov, 2004, p. 96), the following Aldous's condition holds; that is, as $i \to \infty$,

$$\left(\widehat{\wp}_{1n_{k_i}}(\rho_i + \delta_i) - \widehat{\wp}_{1n_{k_i}}(\rho_i),\ \widehat{\wp}_{2n_{k_i}}(\rho_i + \delta_i) - \widehat{\wp}_{2n_{k_i}}(\rho_i)\right)$$
$$\stackrel{d}{=} \left(\widehat{\wp}_{1n_{k_i}}(\delta_i), \widehat{\wp}_{2n_{k_i}}(\delta_i)\right) \xrightarrow{\widetilde{\mathbb{P}}} 0$$

because $\widetilde{\mathbb{P}}(\widehat{N}_{1n_{k_i}}(\delta_i) = 0) = \widetilde{\mathbb{P}}(\widehat{N}_{2n_{k_i}}(\delta_i) = 0) = \exp(-\delta_i \pi_{n_{k_i}}) \to 1$ for each sequence $\{\rho_i, \delta_i\}$, wherever $\{\delta_i\}$ is a sequence of positive numbers converging to zero and $\{\rho_i\}$ is a sequence of stopping times (defined on $\widetilde{\Omega}$) taking values in $[0, T]$ for each fixed $T > 0$. (The stopping time property means that the event $\{\rho_i \leq t\}$ should belong to the σ-field generated by the random variables $(\widehat{\wp}_{1n_{k_i}}(z), \widehat{\wp}_{2n_{k_i}}(z))$ for $0 \leq z \leq t$. See Pollard, 1984, p. 133.)

(b) *Convergence of finite-dimensional distributions.* For any $0 < s_1 < \cdots < s_k$, the characteristic function of $(\widehat{\wp}_{1n_{k_i}}(s_1), \widehat{\wp}_{2n_{k_i}}(s_1), \widehat{\wp}_{1n_{k_i}}(s_2), \widehat{\wp}_{2n_{k_i}}(s_2), ..., \widehat{\wp}_{1n_{k_i}}(s_k), \widehat{\wp}_{2n_{k_i}}(s_k))$ is

$$\begin{aligned}
\widehat{\varphi}_{n_{k_i}}(u_1,...,u_k;v_1,...,v_k) &= \widetilde{\mathbb{E}}\left\{\exp\left[i\left(\sum_{l=1}^{k} u_l \widehat{\wp}_{1n_{k_i}}(s_l) + \sum_{l=1}^{k} v_l \widehat{\wp}_{2n_{k_i}}(s_l)\right)\right]\right\} \\
&= \prod_{l=1}^{k} \exp\left\{-\pi_{n_{k_i}}(s_l - s_{l-1}) \int \left[1-\exp(ia_l x)\right] \right. \\
&\qquad \left. \times d\widehat{F}^{(m)}_{1n_{k_i}}(x)(\omega)\right\} \\
&\quad \times \prod_{l=1}^{k} \exp\left\{-\pi_{n_{k_i}}(s_l - s_{l-1}) \int \left[1-\exp(ib_l x)\right] \right. \\
&\qquad \left. \times d\widehat{F}^{(m)}_{2n_{k_i}}(x)(\omega)\right\},
\end{aligned}$$

where $a_l = u_k + \cdots + u_l$ and $b_l = v_k + \cdots + v_l$. Because $1-\exp(ia_l x)$ and $1-\exp(ib_l x)$ are bounded continuous functions, by (9.7) and Theorem 3.2.3 in Durrett (2010), we have that

$$\begin{aligned}
\widehat{\varphi}_{n_{k_i}}(u_1,...,u_k;v_1,...,v_k) \to &\prod_{l=1}^{k} \exp\left\{-\pi(r_0)(s_l - s_{l-1})\int\left[1-\exp(ia_l x)\right] \right. \\
&\qquad \left. \times dF^{(m)}_1(x|r_0)\right\} \\
&\times \prod_{l=1}^{k} \exp\left\{-\pi(r_0)(s_l - s_{l-1})\int\left[1-\exp(ib_l x)\right] \right. \\
&\qquad \left. \times dF^{(m)}_2(x|r_0)\right\},
\end{aligned}$$

which is the characteristic function of $(\wp^{(m)}_1(s_1), \wp^{(m)}_2(s_1), \wp^{(m)}_1(s_2), \wp^{(m)}_2(s_2), ..., \wp^{(m)}_1(s_k), \wp^{(m)}_2(s_k))$, where

$$\wp^{(m)}_1(z) = \sum_{k=1}^{N_1(z)} Y_k \quad \text{and} \quad \wp^{(m)}_2(z) = \sum_{k=1}^{N_2(z)} Z_k.$$

By Theorem 16 in Pollard (1984, p. 134), it follows that $(\widehat{\wp}_{1n_{k_i}}(\tilde{\omega}, z;\omega), \widehat{\wp}_{2n_{k_i}}(\tilde{\omega}, z;\omega))$ converges weakly to $(\wp^{(m)}_1(z), \wp^{(m)}_2(z))$ in $\mathbb{D}^2[0,\infty)$ for each fixed $\omega \in \mathcal{A}$. Thus, $\widehat{\wp}^{(m)}_{n_{k_i}}(\omega) \Longrightarrow \wp^{(m)}$ in $\mathbb{D}(\mathbb{R})$ as $n \to \infty$ for each fixed $\omega \in \mathcal{A}$. Regard $\widehat{\wp}^{(m)}_n$ and $\wp^{(m)}$ as random elements and denote by $\widehat{P}^{(m)}_n$ and $P^{(m)}$ the induced probability measures, respectively. Let $d_p(\cdot,\cdot)$ be the Prohorov metric. See Billingsley (1999, p. 72). Because weak convergence is equivalent to d_p-convergence (see Billingsley, 1999, Thm. 6.8,

p. 73), then, for each fixed $\omega \in \mathcal{A}$, $d_p(\widehat{P}_{n_{k_j}}^{(m)}, P^{(m)})(\omega) \to 0$. That is, $d_p(\widehat{P}_{n_{k_j}}^{(m)}, P^{(m)}) \to 0$, $\mathbb{P}$-a.s. Hence,

$$d_p\left(\widehat{P}_n^{(m)}, P^{(m)}\right) \to 0, \quad \text{in probability (with respect to } \mathbb{P}\text{)}.$$

Therefore, in probability (with respect to $\mathbb{P}$), $\widehat{\wp}_n^{(m)} \Longrightarrow \wp^{(m)}$ by Theorem 6.8 in Billingsley (1999, p. 73). Furthermore, because the jump distribution $F_k^{(m)}(x|r_0)$ in $\wp^{(m)}$ converges to $F_k(x|r_0)$ as $m \to \infty$, applying Theorem 16 in Pollard (1984, p. 134) again, we can get $\wp^{(m)}(z) \Longrightarrow \wp(z)$ in $\mathbb{D}(\mathbb{R})$ as $m \to \infty$. By Theorem 3.1 in Seijo and Sen (2011), in probability, $\widehat{M}_n^{(m)}$ converges weakly to M_- conditionally on $\mathfrak{X}_n$. ■

REFERENCES

An, H.Z. & F.C. Huang (1996) The geometrical ergodicity of nonlinear autoregressive models. *Statistica Sinica* 6, 943–956.

Billingsley, P. (1999) *Convergence of Probability Measures*, 2nd ed. Wiley.

Brockwell, P.J., J. Liu, & R.L. Tweedie (1992) On the existence of stationary threshold autoregressive moving-average processes. *Journal of Time Series Analysis* 13, 95–107.

Caner, M. (2002) A note on least absolute deviation estimation of a threshold model. *Econometric Theory* 18, 800–814.

Chan, K.S. (1993) Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model. *Annals of Statistics* 21, 520–533.

Chan, K.S., J.D. Petruccelli, H. Tong, & S.W. Woolford (1985) A multiple-threshold AR(1) model. *Journal of Applied Probability* 22, 267–279.

Chan, K.S. & H. Tong (1985) On the use of the deterministic Lyapunov function for the ergodicity of stochastic difference equations. *Advances in Applied Probability* 17, 666–678.

Chan, K.S. & R.S. Tsay (1998) Limiting properties of the least squares estimator of a continuous threshold autoregressive model. *Biometrika* 85, 413–426.

Chen, R. & R.S. Tsay (1991) On the ergodicity of TAR(1) processes. *Annals of Applied Probability* 1, 613–634.

Cline, D.B.H. & H.M.H. Pu (2004) Stability and the Lyapounov exponent of threshold AR-ARCH models. *Annals of Applied Probability* 14, 1920–1949.

Cont, R. & P. Tankov (2004) *Financial Modelling with Jump Processes*. Chapman and Hall.

Durrett, R. (2010) *Probability: Theory and Examples*, 4th ed. Duxbury.

Gooijer, J.D. (1998) On threshold moving-average models. *Journal of Time Series Analysis* 19, 1–18.

Hansen, B.E. (1997) Inference in TAR models. *Studies in Nonlinear Dynamics and Econometrics* 2, 1–14.

Hansen, B.E. (1999) Threshold effects in non-dynamic panels: Estimation, testing, and inference. *Journal of Econometrics* 93, 345–368.

Hansen, B.E. (2000) Sample splitting and threshold estimation. *Econometrica* 68, 575–603.

Hjort, N.L. & M.C. Jones (1996) Better Rules of Thumb for Choosing Bandwidth in Density Estimation. Statistical research report, Department of Mathematics, University of Oslo, Oslo, Norway.

Huber, P.J. (1967) The behavior of maximum likelihood estimates under nonstandard conditions. In L.M. Le Cam & J. Neyman (eds.), *Proceedings of the Fifth Berkeley Symposium on Probability and Statistics*, vol. I, pp. 221–233. University of California Press.

Kapetanios, G. (2001) Model selection in threshold models. *Journal of Time Series Analysis* 22, 733–754.

Kushner, H.J. (1984) *Approximation and Weak Convergence Methods for Random Processes, with Applications to Stochastic Systems Theory.* MIT.

Li, C.W. & W.K. Li (1996) On a double threshold autoregressive heteroscedastic time series model. *Journal of Applied Econometrics* 11, 253–274.

Li, D., S. Ling, & W.K. Li (2010) Least Squares Estimation of the Threshold Moving-Average Model. Working paper, HKUST.

Li, D., S. Ling, & H. Tong (2012) On moving-average models with feedback. *Bernoulli* 18, 735–745.

Li, G.D. & W.K. Li (2008) Testing for threshold moving average with conditional heteroscedasticity. *Statistica Sinica* 18, 647–665.

Li, W.K. & K. Lam (1995) Modeling asymmetry in stock returns by a threshold ARCH model. *Statistician* 44, 333–341.

Ling, S. (1999) On the probabilistic properties of a double threshold ARMA conditional heteroskedastic model. *Journal of Applied Probability* 36, 688–705.

Ling, S. & H. Tong (2005) Testing for a linear MA model against threshold MA models. *Annals of Statistics* 33, 2529–2552.

Ling, S., H. Tong, & D. Li (2007) Ergodicity and invertibility of threshold moving-average models. *Bernoulli* 13, 161–168.

Liu, J., W.K. Li, & C.W. Li (1997) On a threshold autoregression with conditional heteroscedastic variances. *Journal of Statistical Planning and Inference* 62, 279–300.

Liu, J. & E. Susko (1992) On strict stationarity and ergodicity of a nonlinear ARMA model. *Journal of Applied Probability* 29, 363–373.

Lu, Z. (1998) On the geometric ergodicity of a non-linear autoregressive model with an autoregressive conditional heteroscedastic term. *Statistica Sinica* 8, 1205–1217.

Petruccelli, J.D. (1986) On the consistency of least squares estimators for a threshold AR(1) model. *Journal of Time Series Analysis* 7, 269–278.

Pollard, D. (1984) *Convergence of Stochastic Processes.* Springer-Verlag.

Qian, L. (1998) On maximum likelihood estimators for a threshold autoregression. *Journal of Statistical Planning and Inference* 75, 21–46.

Seijo, E. & B. Sen (2011) A continuous mapping theorem for the smallest argmax functional. *Electronic Journal of Statistics* 5, 421–439.

Silverman, B. (1978) Weak and strong uniform consistency of the kernel estimate of a density and its derivatives. *Annals of Statistics* 6, 177–184.

So, M.K.P., W.K. Li, & K. Lam (2002) A threshold stochastic volatility model. *Journal of Forecasting* 21, 473–500.

Tiao, G.C. & R.S. Tsay (1994) Some advances in nonlinear and adaptive modeling in time series. *Journal of Forecasting* 13, 109–131.

Tong, H. (1978) On a threshold model. In C.H. Chen (ed.), *Pattern Recognition and Signal Processing*, pp. 575–586. Sijthoff and Noordhoff.

Tong, H. (1990) *Non-linear Time Series: A Dynamical System Approach*. Oxford University Press.

Tong, H. (2011) Threshold models in time series analysis—30 years on. *Statistics and Its Interface* 4, 107–118.

Tong, H. & K.S. Lim (1980) Threshold autoregression, limit cycles and cyclical data (with discussion). *Journal of the Royal Statistical Society, Series B* 57, 245–292.

Tsay, R.S. (1998) Testing and modeling multivariate threshold models. *Journal of the American Statistical Association* 93, 1188–1202.

Yadav, P.K., P.F. Pope, & K. Paudyal (1994) Threshold autoregressive modeling in finance: The price differences of equivalent assets. *Mathematical Finance* 4, 205–221.

APPENDIX: Weak Convergence of A Pure Jump Process

Let $\{X_k^\varepsilon, k \geq 0\}$, indexed by ε, denote a discrete parameter process generated by

$$X_{k+1}^\varepsilon = X_k^\varepsilon + J_{k+1}^\varepsilon,$$

where the initial value is X_0^ε and $\{J_k^\varepsilon, k \geq 1\}$ is a sequence of jumps. Define the piecewise constant *interpolated process* $x^\varepsilon(t)$ for $t \in [0,1]$ by

$$x^\varepsilon(t) = X_j^\varepsilon, \qquad t \in \left[j\varepsilon, (j+1)\varepsilon\right) \quad \text{for } j = 0, 1, ..., [1/\varepsilon] - 1,$$

and

$$x^\varepsilon(t) = X_{[1/\varepsilon]}^\varepsilon, \qquad t \in \left[[1/\varepsilon]\varepsilon,\ 1\right],$$

where $[1/\varepsilon]$ denotes the integer part of $1/\varepsilon$. What we need is the weak convergence of the interpolated sequence $\{x^\varepsilon(\cdot)\}$. When the limiting process of $\{x^\varepsilon(\cdot)\}$ is an ordinary differential equation or diffusion process, Kushner (1984) gives a detailed and rigorous demonstration through two different methods: the *perturbed test function* method and the *direct-averaging* method. However, when the limit is a pure jump process with J_m^ε being a Markov chain, only an outline is presented. Here, we generalize his result for J_m^ε being measurable in terms of $\mathcal{G}_m = \sigma\{X_i^\varepsilon, i \leq m\}$. Clearly, this result is of interest by itself and can be applied to many other nonlinear time series models. Let $\widehat{C}_0^2$ be a space of functions with compact support and continuous second derivative and let $\mathbb{P}_m^\varepsilon$ and $\mathbb{E}_m^\varepsilon$ be the conditional probability and conditional expectation on $\mathcal{G}_m$, respectively. We first give the following assumptions.

Assumption A.1. For each $\varepsilon > 0$, $\{J_k^\varepsilon\}$ is strictly stationary, and there exists a constant $\lambda \in (0, +\infty)$ such that

$$\lim_{\varepsilon \to 0} \lim_{m \to \infty} \mathbb{P}_k^\varepsilon \left(J_m^\varepsilon \neq 0\right)/\varepsilon = \lambda.$$

Assumption A.2. There exists a random variable U such that $\mathbb{P}(J_k^\varepsilon \in B | J_k^\varepsilon \neq 0) \to \mathbb{P}(U \in B)$ as $\varepsilon \to 0$ for any Borel set $B \in \mathscr{B}(\mathbb{R})$.

Assumption A.3. For any $f \in \widehat{C}_0^2$ and x is a scalar,

$$\lim_{\varepsilon \to 0} \lim_{m \to \infty} \mathbb{E}_k^\varepsilon \left\{f\left(x + J_m^\varepsilon\right) - f(x) | J_m^\varepsilon \neq 0\right\} = \mathbb{E}\{f(x+U) - f(x)\}.$$

Assumption A.4. There is a positive $M < \infty$ such that $|J_k^\varepsilon| \leq M$ for each $k \geq 1$.

Assumptions A.1 and A.2 characterize the jump rate and the distribution of the jump size in the limiting process, respectively. Assumption A.3 is a sufficient condition for the average used in the *direct-averaging* method. Assumption A.4 requires the jumps to be bounded. This is a technical condition. In most applications, the jumps are not bounded in general. We can use the truncated technique to deal with the jumps and consider the truncated process. For some details, see the proof of Theorem 2.3. Based on the preceding assumptions, we have the following theorem.

THEOREM A.1. *Suppose Assumptions A.1–A.4 hold. If* $X_0^\varepsilon \Longrightarrow x_0$, *then* $x^\varepsilon(t) \Longrightarrow x(t)$ *in* $D[0,1]$ *and* $x(t) = J(t) + x_0$, *where* $J(t)$ *is a CPP with jump rate* λ *and jump distribution* $\mathbb{Q}(\cdot)$ *induced by* U *at time* t *and* $J(0) = 0$.

Proof. Let n_ε be an integer satisfying $n_\varepsilon \to \infty$ and $\delta_\varepsilon = \varepsilon n_\varepsilon \to 0$ as $\varepsilon \to 0$. For any fixed function $f(\cdot) \in \widehat{C}_0^2$, define the piecewise constant function

$$\widetilde{A}^\varepsilon f(t) = \frac{1}{\varepsilon n_\varepsilon} \sum_{j=ln_\varepsilon}^{(l+1)n_\varepsilon - 1} \mathbb{E}_{ln_\varepsilon}^\varepsilon \left(f\left(X_{j+1}^\varepsilon\right) - f\left(X_j^\varepsilon\right)\right)$$

for $t \in [l\delta_\varepsilon, (l+1)\delta_\varepsilon)$. Clearly, it follows that

$$\mathbb{E}_k^\varepsilon \left\{ f\left(X_{m+1}^\varepsilon\right) - f\left(X_m^\varepsilon\right)\right\} = \mathbb{P}_k^\varepsilon\left(J_m^\varepsilon \neq 0\right) \mathbb{E}_k^\varepsilon \left\{ f\left(X_m^\varepsilon + J_m^\varepsilon\right) - f\left(X_m^\varepsilon\right) \middle| J_m^\varepsilon \neq 0\right\}.$$

By Assumption A.1, we have

$$\begin{aligned}\widetilde{A}^\varepsilon f(t) &= \frac{1}{\varepsilon n_\varepsilon} \sum_{j=ln_\varepsilon}^{(l+1)n_\varepsilon - 1} \mathbb{E}_{ln_\varepsilon}^\varepsilon \left\{ f\left(X_{j+1}^\varepsilon\right) - f\left(X_j^\varepsilon\right)\right\} \\ &= \frac{\lambda}{n_\varepsilon} \sum_{j=ln_\varepsilon}^{(l+1)n_\varepsilon - 1} \mathbb{E}_{ln_\varepsilon}^\varepsilon \left\{ f\left(X_j^\varepsilon + J_j^\varepsilon\right) - f\left(X_j^\varepsilon\right) \middle| J_j^\varepsilon \neq 0\right\} + o(1).\end{aligned}$$

Let $\widehat{A}^\varepsilon(t) = \int_0^t \widetilde{A}^\varepsilon f(s) ds$. By the boundedness of λ and J_m^ε in Assumptions A.1 and A.4, it follows that $\{(x^\varepsilon(t), \widehat{A}^\varepsilon(t))\}$ is tight in $D^2[0,1]$. In fact, for the tightness of $x^\varepsilon(t)$, see Kushner (1984, the last paragraph on p. 32). The tightness of $\widehat{A}^\varepsilon(t)$ is implied by the boundness of $\widetilde{A}^\varepsilon f(t)$ due to $f \in \widehat{C}_0^2$. Because it is sufficient to work with an arbitrary weakly convergent subsequence also indexed by ε, without loss of generality, suppose that $(x^\varepsilon(t), \widehat{A}^\varepsilon(t)) \Longrightarrow (x(t), \widehat{A}(t))$ in $D^2[0,1]$. By means of the *Skorokhod embedding theorem* in Kushner (1984, p. 29), we assume that $(x^\varepsilon(t), \widehat{A}^\varepsilon(t))$ converges to $(x(t), \widehat{A}(t))$ a.s.

Let $\mathcal{C} = \{s \in [0,1] : x(t)$ is continuous at $s\}$. Then for any $s \in \mathcal{C}$, there exists an integer l_ε such that $s \in [l_\varepsilon \delta_\varepsilon, (l_\varepsilon + 1)\delta_\varepsilon)$. Let $m_\varepsilon = l_\varepsilon n_\varepsilon$. Then, for $f(\cdot) \in \widehat{C}_0^2$, by Assumptions A.2 and A.3, it follows that

$$\begin{aligned}\widetilde{A}^\varepsilon f(s) = \lambda &\left\{ \frac{1}{n_\varepsilon} \sum_{j=m_\varepsilon}^{m_\varepsilon + n_\varepsilon - 1} \mathbb{E}_{l_\varepsilon n_\varepsilon}^\varepsilon \left\{ f\left(X_j^\varepsilon + J_j^\varepsilon\right) - f\left(x(s) + J_j^\varepsilon\right) \middle| J_j^\varepsilon \neq 0\right\}\right\} \\ &+ \lambda \left\{ \frac{1}{n_\varepsilon} \sum_{j=m_\varepsilon}^{m_\varepsilon + n_\varepsilon - 1} \mathbb{E}_{l_\varepsilon n_\varepsilon}^\varepsilon \left\{ f(x(s)) - f\left(X_j^\varepsilon\right) \middle| J_j^\varepsilon \neq 0\right\}\right\} \\ &+ \lambda \left\{ \frac{1}{n_\varepsilon} \sum_{j=m_\varepsilon}^{m_\varepsilon + n_\varepsilon - 1} \mathbb{E}_{l_\varepsilon n_\varepsilon}^\varepsilon \left\{ f\left(x(s) + J_j^\varepsilon\right) - f(x(s)) \middle| J_j^\varepsilon \neq 0\right\}\right\} + o(1) \\ &\to \lambda \mathbb{E}\left\{ f(x(s) + U) - f(x(s))\right\} \\ &= \lambda \int \left[f(x(s) + u) - f(x(s))\right] \mathbb{Q}(du) \equiv Af(x(s)).\end{aligned}$$

Thus, $\widehat{A}(t) = \int_0^t Af(x(s))ds$. For arbitrary k, t, s with $s_1 < s_2 < \cdots < s_k < t < t+s \leq T$ and any bounded and continuous function $g(\cdot)$, by Taylor's expansion, it follows that

$$\mathbb{E}\left\{g\left(x^{\varepsilon}(s_j), j \leq m\right) \times \left[f\left(x^{\varepsilon}(t+s)\right) - f\left(x^{\varepsilon}(t)\right) - \int_t^{t+s} \widetilde{A}^{\varepsilon} f(u)du\right]\right\} = \triangle_{\varepsilon},$$

where $\triangle_{\varepsilon} \to 0$ as $\varepsilon \to 0$. Hence,

$$\mathbb{E}\left\{g\left(x(s_j), j \leq k\right)\left[f(x(t+s)) - f(x(t)) - \left(\widehat{A}(t+s) - \widehat{A}(t)\right)\right]\right\} = 0,$$

which implies that $x(\cdot)$ solves the martingale problem for the operator A and the initial condition x_0. That is,

$$f(x(t)) - \int_0^t Af(x(s))ds \quad \text{is a martingale for the operator } A.$$

Then $x^{\varepsilon}(t) \Longrightarrow x(t) = J(t) + x_0$ in $D[0,1]$, where $J(t)$ is a CPP with jump rate λ and the jump distribution $\mathbb{Q}(\cdot)$ and $J(0) = 0$. ■